

Should spycatchers write books?

Last week's decision by the House of Lords that British newspapers can now refer to Peter Wright's infamous book is not a victory for press freedom but a pointer to illiberal restraints ahead.

FOR the past three years, the international readership of *Nature* has been arbitrarily partitioned into two sets.

One, the larger, has been able to read without let or hindrance a book called *Spycatcher*, written by a one-time British counter-spy called Peter Wright now living in Australia. The other group is that of readers living in Britain, and whose interest in what Mr Wright says is likely to be the greater. They have been denied access to the book for the past two years as a consequence of a series of injunctions against British newspapers granted the British government by the British courts.

Technically, it would no doubt have been possible for this journal to have satisfied the bulk of its readers' interests in *Spycatcher* by discussing the book only in copies circulating elsewhere than Britain, perhaps in the copies of *Nature* manufactured in the United States, Japan and China, which would have been a nonsense. Now, the British House of Lords (which functions imperfectly as Britain's supreme court) has decided that the injunctions must be lifted.

Not before time. In its pursuit of Mr Wright and *Spycatcher* through the courts of Australia, Hong Kong, New Zealand and Britain, the British government has made fools of itself and its constituents. So much could have been (and was) clear from the start. For while there is no doubt that the book should never have been published and that its appearance anywhere is both damaging to British interests and a breach of the duty which every employee owes an employer to keep confidential matters confidential, it has always been clear that the disputed book would have found its way into print somewhere and from there to almost everywhere. For Mr Wright had taken the precaution of living outside the jurisdiction of the British courts. Even if the Australian courts had granted the British government the injunction against publication for which it asked, Mr Wright would have been able to publish his book in the United States with the protection of the First Amendment of the Constitution.

The British government now says that its zeal in the fruitless prosecution of *Spycatcher* must be understood as an act of principle which, if not taken, would have been regarded by other spies with literary aspirations to neglect their proper obligations. That argument would carry more weight if the publicity of the court cases had not had not enormously enriched Mr Wright.

The British government's case against Mr Wright and his book is not in dispute. The three sensational allegations towards the end of the book (that there had been a plot to undermine a former prime minister, Mr Harold Wilson, by disinformation, that there has earlier been a plot to assassinate the Egyptian leader, Colonel Nasser, and that a former head of the counter-espionage services, Sir Roger Hollis, was himself a double agent) would be damaging if true (for which reason there is a strong case for a full-blooded enquiry into them). But these are the book's insubstantial sensation.

Bugging

The incontrovertible damage the book does is Mr Wright's description of the working of the British counter-espionage services. There is, for example, a meticulous account of the bugging of the Soviet consulate in London that cannot but be

true. Mr Wright, himself employed during the 1950s as the first member of the counter-espionage services with responsibility for research, says enough about the technology of detecting radio receivers in operation to give emulators a useful spring-board.

Wright's own complaint that the counter-spies too often behave like amateurs is well supported by his data. Almost in passing, the book drops the names of people until recently in public life who must be supposed to have been disloyal to their employer, the British government. With signs that the dust may now settle, the British Establishment is fond of saying that *Spycatcher* is only a second-rate book, but nothing could be further from the truth. It is a rattling good yarn which, in most places, rings of the truth.

The same, unfortunately, cannot be said of the opinion of the Law Lords, decisive (for the newspapers and publishers immediately affected) and influential (for the future of British legislation on secrecy) though it is bound to be. Not that anybody will seriously disagree with the opinion that Mr Wright would not have been able to publish his book in Britain; the government, the Law Lords say, would have easily won an injunction against it. But would not publication also, and more simply, have been prevented by the Official Secrets Act, still in force despite the government's declared intention earlier this year that the act will be reformed. (The present act makes all unauthorized disclosure of official secrets a criminal offence; while the proposed amendment would define categories of protected information more narrowly, it is intended that there should be no legal defence against a charge of disclosing information of the kind Mr Wright has been retailing.)

Precedents

Instead, the House of Lords seeks to base its conclusion that one newspaper (*The Sunday Times*) behaved reprehensibly in seeking to serialize the book (one installment appeared) on the more general but much more dubious law of confidentiality. The Law Lords rely on precedents such as that in which a duke was restrained from telling the secrets of the marriage-bed he had shared with his estranged wife and on the hypothetical example of a charitable donor who wishes to remain anonymous (and whose right to do so are supposed to be easily enforceable on third parties in the British courts). This amounts to a declaration that interested parties are free to decide among themselves what shall be confidential, and that publishers will then have a general duty to respect these confidences.

Such a principle, if generally adopted, would plainly offend against the public (as distinct from the governmental) interest in respect of Mr Wright's more sensational (but still unproven) allegations. While the purpose of this argument based on confidentiality is to purport to show that publishers are bound by the private decisions of persons and individuals, so as to conclude that a British newspaper printed an extract from *Spycatcher* would be guilty of a breach of confidence, however important the material, it is especially disturbing that this argument is likely to be used in the arguments ahead about the drafting of the amended Official Secrets Act. ▶



It is especially disturbing that Lord Griffith (the judge who happened to be the chairman of the British Government's commission on the security services) goes so far as to say that editors faced with the question whether an official secret should be published, after ascertaining that their information is correct, should then consult the government before publication.

Newspapers, in other words, possessing information that may well be embarrassing to governments should, if they believe the information to be true, offer the government an opportunity to restrain them from publication. That would be like saying that President Ronald Reagan's administration should have the right to decide whether US newspapers should publish information about the sale of arms to Iran. In logic, it should not be hard to dissuade the British government from following this course in drafting its new legislation, but the British animus against the press is now so great that logic may be insufficient.

Those are some of the reasons why last week's judgement is not nearly the vindication of a free press it has been hailed as being. In the event, the Law Lords conclude that the injunctions against British newspapers should be lifted because *Spycatcher*, as a consequence of the ingenuity of international book-publishers, is now readily available in Britain. That is mere commonsense, so that it is especially ironical that the proposed amendment of British secrets law will (if the government has its way) specifically exclude as a defence against prosecution proofs that the disputed information is already generally known. This plan to make Britain into a kind of island of sovereign secrets is especially ridiculous when more and more of the channels of communication (even *Nature*) are increasingly international in character.

Lessons

How else might the British government have set about the job of dealing with *Spycatcher*? The first need is to acknowledge that the British had every cause to be outraged that Mr Wright had chosen to breach his fiduciary obligation to his employers, and to spill all those fascinating beans. The first lesson is to make sure that all similarly placed people are aware of their obligations and that there are mechanisms by which they can be enjoined, not merely exhorted, to comply. (If the British Treasury had been less niggardly about Mr Wright's pension, he might never have written his book, but, even if he had, the government would have had a hold on him.) The second is to ensure that the British security services themselves are less ready to use the leak of confidential information as a means of bolstering their own cause. Wright's book shows compellingly that spying, while dangerous, is also tedious and demeaning, but there is a quite different vein of British spy literature clearly inspired by secret agents to glamorize their own work. (One successful author is both an ex-spy and a Member of Parliament.) In the case of *Spycatcher*, the British government's case in the Australian courts was weakened by its apparent indifference to the earlier publication by Mr Chapman Pincher dealing with Wright's allegations against Sir Christopher Hollis. (Lord Rothschild's role in that affair needs to be clarified.)

There is also an urgent need of action internationally. One of the Law Lords, Lord Keith, argues that it might be possible to silence talkative spies who have escaped the jurisdiction of their employer governments by means of an international agreement; no doubt something could be done. But cases like that of Peter Wright are small beer compared to those in which major spies defect from one side to another whereupon the often startled host will thoughtlessly and pointlessly make loud propaganda about what seems to them success. Would not an agreement on self-restraint be better? And while, at least in the relationship between major powers, the interests of the spies is turning from military to political matters, is it entirely impossible that all embassy spies should be registered as such? That way, there would be many fewer occasions on which groups of diplomats are sent packing at short notice. □

Nuclear options ahead

Safe nuclear reactors should be the goal. Abolition would be shortsighted.

THE need to make materials for nuclear weapons programmes may not be an unmitigated nuisance, but may also be an opportunity. The nuclear industry in the United States is in a curious period of transition. For more than a decade, the electricity utilities have been glad to manufacture their electricity from sources other than uranium, and have been openly saying as much, cheered on by small armies of protesters. Now it seems as if the US Department of Energy will be running into troubles similar in kind because of its continuing responsibility for the production of the ingredients from which nuclear weapons are built. During the past few weeks, there has been a spate of complaint that many of the department's production plants, scattered over fifteen sites in the continental United States, have either become unsafe or have been so for several years. Now, no doubt wishing to anticipate what its critics will be saying, the department has put out what purports to be an unvarnished account of what is really going on at the weapons plants.

Too many of the plants are now sources of radioactive contamination that would long since have been closed if their purpose were civilian, not military. From Hanford in the state of Washington to the Savannah River Plant (see page 662) in South Carolina, production plants first built between thirty and forty years ago have been found to be leaking radioactivity. It is natural that the department's officials should share the alarm expressed by the environmentalists. At the same time, it is only fair to acknowledge that even if the department's production plants had been well-designed and built, the passage of time would by now be taking its toll of them.

The truth is that the United States cannot do without them. Even if there should be a bilateral agreement reducing the numbers of strategic warheads deployed by the United States and the Soviet Union, it will still be necessary that the tritium content of the weapons that remain should from time to time be replenished. Moreover, facilities for the production of enriched uranium and plutonium will still be needed even if the stockpile of warheads should be marvellously reduced by bilateral agreement; there will always be a risk that an agreement might fall apart. Yet the coincidence of the ageing of the weapons production plants and improved prospects for an agreement on strategic weapons should suggest to the Department of Energy that the need, now, is not for the prosecution of a new construction programme but for a cool consideration of a production plan for the first half of the next century. A calculation of the effort that might be saved if there were a durable arms control agreement would be entirely appropriate.

The department also seems to have other ambitions. A few months ago it became known that the department would like to save on the costs of military production by means of dual-purpose reactors able to produce heat energy for the civil economy as well as to manufacture weapons materials. Although this flies in the face of the principle on which the department was founded, that there should be a sharp division between military and civilian affairs in atomic energy, there is no reason why the idea should not be given a sympathetic hearing.

In its public reaction against nuclear energy, the US population has driven itself into a corner more extreme than necessary. Should not the objectors give up crying that nuclear power is an abomination, and start looking for safe reactors instead? Those now in service in the United States, are cooled with water at high-pressure because that is what the nuclear submarine programme, begun in the 1950s, required. A modest exploration of better designs would surely help guard against the painful dilemma that will arise when the United States is as worried about greenhouse gases as about radioactivity. □

Surprising new ozone data from NASA satellite

- Unusual climate conditions
- Prognosis remains poor

Washington

THIS year's Antarctic ozone hole will be much shallower than last year's. Measurements from the Nimbus-7 satellite, announced last week by NASA (National Aeronautics and Space Administration), revealed only a 15 per cent drop in ozone concentration during September, compared to a 50 per cent drop during the same month last year. Ozone depletion follows a two-year cycle, with greater losses occurring during odd-numbered years, but the 1988 ozone hole promises to be the least since 1982, marking the first time the general pattern of deterioration has not been followed.

This year's relatively small ozone hole is related to the general meteorology of the Southern Hemisphere. The trend of ozone loss follows the quasi-biennial oscillation (QBO), a two-year cycle in the global atmospheric circulation pattern. The QBO affects the size and stability of the polar vortex, the isolated region of Antarctic air inside which pollutants that can destroy ozone accumulate. According to

Mark Schoeberl, of the NASA/Goddard Space Flight Center, the polar vortex has been sporadically disrupted this year by unusual climatic conditions, and air from the mid-latitudes has flushed out some of the ozone-destroying chlorinated gases. In addition, the stratospheric air is somewhat warmer than usual, so that there are fewer icy clouds, within which most of the ozone-destroying reactions occur.

The sensitivity of the ozone depletion to rather small variations in the QBO is surprising, and underlines the many unanswered scientific questions concerning the relation of the Antarctic air to the rest of the atmosphere. Despite these uncertainties, the outlook for future years remains bad. In 1987, ozone was almost completely destroyed within the polar vortex, so that ozone concentrations cannot fall much below the minimum levels they reached last year. But Schoeberl says that as pollutant levels build up, ozone destruction will proceed more rapidly, and the Antarctic hole will appear earlier in the season.

David Lindley

Britain's ozone research lags a long way behind the rest

London

STRATOSPHERIC ozone research in the United Kingdom is underfunded and uncoordinated. Funds should be doubled and a national committee should be set up to establish and support a national programme of research. Those are the main conclusions and recommendations of the second report produced by the United Kingdom Stratospheric Ozone Research Group, which advises the government on policy.

Dr John Pyle of the University of Cambridge, chairman of the group, says that the current budget of £2 million is inadequate to carry out vital basic and strategic research. As a result this country is tagging on to the research programmes of others and so cannot determine the direction of its research. "We need to have a programme of our own," he said.

Sources are fragmented and uncoordinated. And some projects are delayed or rejected because it is unclear which research council is responsible for that research. The delineation of the responsibilities of the Natural Environment Research Council and the Science and Engineering Research Council need to be

clarified, says the report.

The group recommends greater co-ordination of stratospheric ozone research in Europe. And says it was "particularly disappointing" that this research was excluded from the current environment programme of the Commission of the European Communities.

In order to redress this, a separate task force is being set up by the commission jointly with the environment ministers of the United Kingdom and Norway. At a meeting in the Hague this week, it will decide how to use its budget of £180,000 to promote and coordinate research within the European Community and countries of the European Free Trade Association. Also at the Hague this week are four international meetings on stratospheric ozone organized by the United Nations Environment Programme. They include a review of current research, discussions on substitutes for ozone-depleting chemicals, and on the harmonization of data on the production, import and export of the offending chemicals, as well as discussions on the implementation of the Montreal Protocol, due to take effect in 1990.

Christine McGourty

Congress passes bill to fight AIDS

Washington

MEASURES aimed at boosting research on AIDS sailed through the US Congress last week in the first comprehensive legislation designed to counter the AIDS epidemic. But other more controversial sections of the bill that guaranteed the confidentiality of results from AIDS tests had to be sacrificed when a handful of conservative senators threatened to block passage of the bill.

The bill is an amalgam of measures introduced in the Senate by Edward Kennedy (Democrat, Massachusetts) and in the House of Representatives by Henry Waxman (Democrat, California) and fought over for almost two years. Most hotly disputed were portions of the bill designed to protect the rights of those suffering from AIDS. Specific clauses outlawing discrimination against persons with AIDS were deleted early on in the debate, but the bill's sponsors won victories when amendments that would have required premarital testing, and also the testing of the entire US prison population, were defeated (see *Nature* 335, 386; 1988).

A final hurdle set up by conservative Senator Jesse Helms (Republican, North Carolina) brought down clauses that strictly guaranteed the confidentiality of AIDS tests results and mandated fines and prison penalties for violators. The bill's supporters argue that protection of confidentiality is essential if those who suspect they may carry the human immunodeficiency virus (HIV) are to come forward for testing without fear of victimization. But rather than see the whole bill defeated, they agreed to amend the bill, allowing doctors to reveal test results to third parties in limited circumstances. Plans to spend \$400 million a year for a period of three years on testing and counselling of people with high risk of exposure to HIV will be cut back to the \$100 million level.

The bill will add a total of 780 new posts to the major federal agencies carrying out research on AIDS, principally the National Institutes of Health, the Centers for Disease Control and the Food and Drug Administration. An extra \$300 million a year for three years will be used to supplement funds for AIDS research provided elsewhere. Specific measures provide for a new international AIDS research data bank, increased collection and analysis of demographic data, and steps to speed up testing procedures for promising drugs and vaccines.

Alun Anderson

Pharmaceuticals trio a popular choice for Nobel awards

London

If this year's Nobel prize in Physiology or Medicine was intended to cover the water-front of drug discovery, it could hardly have been better chosen. Sir James Black, the 62-year-old British winner, is credited with the discovery of drugs that block β -adrenergic receptors and histamine H₂ receptors, while the pair of US winners, 70-year old Gertrude Elion and 83-year old George Hitchings, discovered inhibitors of purine and pyrimidine metabolism, which have been of value in the treatment of leukaemia, malaria, gout, organ transplantation and bacterial infections.

Black spent 12 years in university posts before switching to the pharmaceutical industry, and has oscillated between the two ever since. In his first industrial post, as senior pharmacologist with Imperial Chemicals Industries, Black together with William Duncan developed propranolol, the first 'beta blocker'. Numerous drugs of this type are used in the treatment of angina and hypertension.

Propranolol was produced in 1964, the year Black and Duncan joined Smith Kline and French Laboratories to apply to histamine receptors the approach that had produced beta blockers. "About 700 compounds have been synthesized and tested", said their 1972 paper that announced the prototype histamine H₂ antagonist, burimamide (*Nature* **236**, 385; 1972). The most successful of the company's subsequent H₂ antagonists has been cimetidine (Tagamet), which is widely used in the treatment of peptic ulcers.

In 1973, Black left industry for the University of London, but by 1977 he was ready to return. Smith, Kline and French beckoned him to the United States, but a last minute bid kept Black in Britain. He was director of therapeutic research for Wellcome Research Laboratories until 1984, since when he has been back in London University at King's College Hospital Hospital Medical School.

Wellcome employed the other two prize winners from the mid-1940s until their semi-retirement. One result is their citation as co-inventors on 18 patents for purine and pyrimidine metabolites. In the words of a colleague, "they defined and delineated with unusual ingenuity the pathways of purine and pyrimidine metabolism".

Hitchings worked on both aspects, while Elion concentrated on the pyrimidines. Among the first drugs that emerged were 6-mercaptopurine and thioguanine, both active against leukaemia, and pyrimethamine (Daraprim) and trimethoprim, both inhibitors of dihydrofolate

reductase and used as antimalarials. Trimethoprim, usually in conjunction with a sulphonamide, is also a much-used antibiotic.

More recently, Hitchings and Elion developed azathioprine (Imuran), the immunosuppressive drug that has probably been used more than any others to decrease the rate of rejection of transplanted organs, and allopurinol, the xanthine oxidase inhibitor that is used in the treat-

ment of gout and kidney stones. While their citation also mentions acyclovir, it was more the anti-metabolite philosophy of Elion and Hitchings than their direct input that produced this recent anti-herpes drug.

While the rumour-mongers were firmly tipping oncogenes or retrovirus, perhaps including the human immunodeficiency virus, as the topics that would provide the winners of this year's prize, the Nobel Committee has opted for "important principles for drug treatment", and has chosen a trio whose recognition was, in the opinion of many pharmacologists, long overdue.

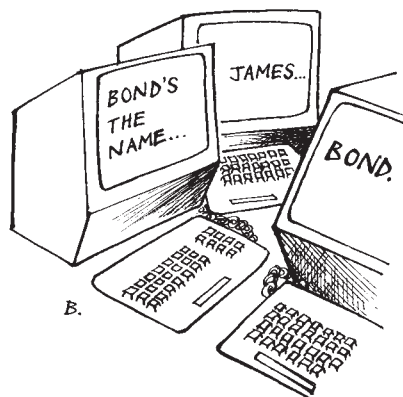
Peter Newmark

Numbers game is of prime importance to spies

Berkeley

By using electronic mail networks to link hundreds of computers an international team of computer scientists has succeeded in factoring a 100-digit number into two large prime factors. This task was until recently thought impossible because of the massive amount of expensive computer time needed — up to half a year on a Cray supercomputer.

The achievement has special signifi-



cance for some professions, because it could make it easier for spies to break code systems. A common form of encryption relies for its security on the difficulty of factoring very large numbers. "People designing keys should be careful", said project co-organizer Arjen Lenstra, of the University of Chicago, "because there is much more computer power around than they realize". The technological advance that put the problem within reach was made in 1985 when Robert Silverman of Mitre Corporation in Bedford, Massachusetts, found a means of breaking up the factorization algorithm so it could be run in parallel, on a network of computers. Though the calculation used no new mathematical approaches, the organizers say it shows the ease of parallel computing through electronic mail.

Mark Manasse, of the Digital Equipment Corporation's Systems Research

Center in Palo Alto, California, said he and Lenstra began the project to explore the power of loosely-coupled distributed processing between machines that "barely know each other". The factorization algorithm was broken down to parts and sent to each of 400 participating computers in Holland, the United States and Australia, which then reported back to Manasse daily, and supplied enough data for him to solve the problem in less than four weeks.

Any computer that uses C language could participate, and the network included micro-VAXs, Suns, and a Cray. The network "didn't spend one single penny for computer time", said Lenstra, because the computers were devoted to the problem only during time when they would otherwise have been idle. While only 30 per cent of the computers working on the 100-digit number were from outside Digital's Palo Alto facility, the outside arm of the project is growing, and Manasse expects to amass 5,000 to 20,000 computers, enabling the network to factor larger numbers.

Lenstra said the experiment may have increased the safe threshold for the popular RSA code system, in which codes based on large numbers of 100 or more digits can be broken by anyone able to factor the number. Loosely coupled computer networks may bring factorization within the reach of would-be code breakers.

Silverman doubts that the DEC experiment has given cryptographers anything to worry about, however. RSA system users have been advised for some time to base their codes on numbers of at least 150 digits, he said, a size that Lenstra and Manasse agree remains unfactorable with currently-known algorithms. Since each addition of three digits doubles the computing power necessary to factor the number, a 150-digit number would require over 100,000 times the computing power for a 100-digit number. Marcia Barinaga

Curien advocates a long-term strategy for civil research

Paris

FRENCH research minister Hubert Curien last week announced the creation of a new body to monitor the nation's science and technology research activities. This "observatory", he says, has become vital given the current levels of government research funding. It is one of several initiatives being taken by the government to prepare a long-term strategy for civil research following what Curien has acknowledged is a 'once only' budget aimed to inject cash into research without a clear long-term plan (*Nature* 334, 556; 1988).

In this latest move, Curien has appointed Pierre Papon, former director of the CNRS (Centre National de la Recherche Scientifique) and lecturer at the Ecole supérieure de physique et de chimie industrielle, to "find participating members for the monitoring body, define its scope and to set out a timetable for its realization". Curien expects Papon to submit his report by the end of December so that the project can start in January 1989.

The new body has already been set a number of specific tasks: to embrace the entire national system of science and technology, to monitor its evolution and to relate French efforts to European and worldwide initiatives. To do this, Curien has promised that the monitoring body will "offer all the guarantees of objectivity necessary for its credibility and will involve representatives of the national research system".

The new government, under a banner of *glasnost*, is keen to show that coherent long-term policies are being formulated contingent upon reports from expert working parties. In the meantime few ministers have been willing to say what these policies are likely to contain, and one minister was even asked to resign when he broke this tacit rule of silence. Health minister Claude Evin is, however, expected to announce shortly the government's plans to combat AIDS using the recommendations of his advisor, Claude Got.

Peter Coles

Competition for Soviet research funds

London

THE distribution of financial support for research in the Soviet Union is to be decided by open competition between laboratories, and a secret ballot by a 'competent commission', academician Gurii Marchuk, president of the Soviet Academy of Sciences, told the official Soviet press agency TASS recently. Under the former system, research funds were handed over to the heads of institutes, who then shared the money out between the various laboratories under

their administration. Under the new system, intended to implement Gorbachev's economic reforms, the heads of laboratories will themselves compete for funds.

In the case of applied research, each laboratory head must submit proposals for solving a particular problem, and the commission will then vote on who is to receive support. For research with no direct applications, a special fund will be established, but grants from this will also be based on competition, Marchuk said.

Marchuk stresses that more must be done in bridging the gap between basic research and industrial applications and urged a 50 per cent increase in allocations for building facilities such as pilot plants. Whether the new financial arrangements will prove workable is open to question. The funding of research on the basis of specific problems was tried in Poland during the 1970s under the auspices of the then minister of science, Dr Sylwester Kaliski, but ran into difficulties when, for example, an institute found it necessary to purchase an expensive item of equipment which would be shared by teams and laboratories working on different topics. This scheme was eventually abandoned on the grounds that it imposed too many bureaucratic constraints.

Whether the Soviet scheme can avoid such difficulties will largely depend on the efforts of the decision-making commission which, some Soviet scientists have suggested, should be empowered to call in foreign scientists as referees. Vera Rich

New Hungarian union

THE new general secretary of Hungary's trade union federation, SZOT, Mr Sandor Nagy, has stated that he "personally acknowledged" the fact of the existence of TDDSZ, the breakaway Democratic Union of Scientific Workers "as a reality". Following talks between representatives of SZOT and TDDSZ last month, Nagy told the daily *Magyar Nemzet* that the existence of TDDSZ was, however, a "sad reality" for the traditional union movement, "since those who operate independently from us imply criticism of official trade union activities". The TDDSZ members broke away from SZOT earlier this year, complaining that their interests, as scientists, could not receive proper attention in a single-union-per-workplace structure. Under the old system, scientists were spread among 18 unions, but constituted only a minority in each. Vera Rich

Mutual friends

THE United States and India have signed a three-year extension to a bilateral Science and Technology Initiative, first established in 1982 by President Reagan and Prime Minister Indira Gandhi. US Science Advisor William Graham met Prime Minister Rajiv Gandhi in New Delhi earlier this month when he praised the joint research programmes, but emphasized the importance of mutual respect for each other's intellectual property rights. J.P.

Polio outbreak

FOLLOWING an outbreak of polio that has claimed ten victims in less than two months, Israeli health ministry officials have announced a nationwide immunization programme. People under 40 years of age in Israel, the West Bank and Gaza will all receive oral polio vaccine boosters as a 'preventive measure'.

Health officials say the outbreak may be linked to polio virus found in sewage samples in seven sites around the country. L. P.

Banking on commerce

THE industrial sector in Britain is losing graduates to finance and commerce according to the latest statistics. The number of students entering banking and insurance increased by 18 per cent in 1986-87, and mathematicians, engineers and technologists entering industry decreased. The number of graduates going into research increased slightly, reversing the downward trend which began in 1983-84. C.McG.

Secret weapons labs

SUSPECTED foreign agents have had access to sensitive information at US weapons laboratories, according to a new report by the General Accounting Office, an investigative arm of Congress. GAO says the Department of Energy has not done an effective job enforcing regulations for screening foreign visitors. In addition to visits from citizens of Warsaw Pact nations and China, there are numerous visits by citizens of other countries considered capable of building nuclear weapons.

DoE officials deny that a serious problem exists and point out that for each of the weapons laboratories there are classified and non-classified areas. J.P.

British export

OXFORD University may soon be setting up a branch in Japan. St Catherine's College is discussing with the Japanese manufacturing company Kobe Steel opening a small college in Japan to provide students there with a general education on life in Britain and prepare them for undertaking studies in this country. The college would take no more than 50 students and provide courses in history, law, economics and mathematics. If all goes well the college will open in 1990. C.McG.

Nuclear enrichment for Japan despite growing worries

Tokyo

JAPAN Nuclear Fuel Industry Co. began construction on Friday of a uranium enrichment plant in northern Japan that will form part of a huge commercial complex to enrich uranium, reprocess spent nuclear fuel, and store low-level radioactive waste. The complex, which is expected to cost one million million yen (\$8,000 million), is intended to reduce Japan's dependence on the US and Europe for nuclear fuel. But construction of the facility is expected to meet strong opposition from the growing anti-nuclear movement in Japan.

The enrichment plant under construction at Rokkasho on the northern tip of the main island of Japan is scheduled to begin operations in 1991. It will initially produce 600 tons SWU (separate work units) of uranium per year and by 2000 this will be increased to 1,500 tons, enough for twelve 1 megawatt reactors or about 30 per cent of Japan's needs.

At present Japan is almost entirely dependent on the US for supplies of enriched uranium and, as a result, reprocessing of Japan's spent fuel is subject to strict US non-proliferation regulations. A reprocessing plant scheduled to open at the Rokkasho site in 1995 will be able to handle a large proportion of Japan's spent nuclear fuel by the turn of the century and this will give Japan greater freedom to handle its nuclear fuel supplies. Vast amounts of spent Japanese nuclear fuel are piled up in reprocessing plants in the UK and France. Shipments of plutonium derived from reprocessing of the fuel are due to begin in the early 1990s and the means of transporting the material has become an issue of great controversy in the US (see right).

Nuclear Fuel Industry Co. has opted to use a conventional centrifugal technique for uranium enrichment at its Rokkasho plant. Laser enrichment techniques are now under development in the US, Japan and Europe and there are fears that the technology at Rokkasho may be outdated by the time it comes into operation.

But competing technology is probably not the biggest of the company's present worries. This year has seen a tremendous surge in growth of the anti-nuclear movement in Japan (*Nature* 333, 199; 1988). The government and electric power companies have spent millions of dollars over the past few months on public relations campaigns. The Nuclear Safety Commission's latest annual report released last week contains for the first time a question and answer section derived from public hearings to reassure the public of the safety of nuclear power. And on Saturday

the first ever drill to practice evacuation of the public was held in the area surrounding a new nuclear power plant in the northern island of Hokkaido.

Chernobyl has slowly awoken urban dwellers to the fact that they may not be immune from the dangers of nuclear accidents at plants located in remote country areas. Takashi Hirose, a writer and anti-nuclear activist, has been particularly influential in cultivating doubts in the minds of the general public. His book entitled *Kiken na hanashi* (a dangerous

story) is a best seller and he lectures to packed audiences around the country. His message is simple. Minor accidents do happen quite often at Japanese nuclear power plants (although much less frequently than in other countries). No-one can guarantee that a major accident will not occur and as the nuclear industry in Japan rapidly expands the chances of an accident increase.

The complex at Rokkasho has become a focal point of the anti-nuclear campaign. Local groups have collected 600,000 signatures opposing construction. Five hundred antinuclear demonstrators protested at the remote Rokkasho site on October 9 and further protests are expected.

David Swinbanks

Sea transport of plutonium approved despite Congress

Washington and Tokyo

JAPAN has won US approval to transport large quantities of weapons-grade plutonium by sea from Europe despite Congressional and Department of Defense fears that a ship would be a tempting target for a terrorist attack. Only a few kilograms of the plutonium are needed to make a nuclear bomb.

On the one past occasion that Japan carried plutonium from France by sea, the ship was escorted by US, French and Japanese naval vessels and was constantly monitored by satellite. But although the new US-Japan agreement recommends that a military escort be provided, it contains a vaguely-worded proviso that could make an unescorted shipment possible.

Only three months ago the United States and Japan came to a different agreement. The "US-Japan Agreement on Peaceful Use of Nuclear Power" granted Japan 30-year blanket approval to reprocess nuclear fuel of US origin, which is subject to strict US nuclear non-proliferation regulations, provided the plutonium derived from reprocessing in Europe was transported back to Japan by air.

The policy turn-around has angered Congress, largely because of the way it was carried out. Rather than asking that the previous agreement be amended to allow sea transport, giving Congress a 90-day review period, the State Department regarded the change as a special case "subsequent arrangement" that would automatically become effective unless Congress took action within fifteen days. The time expired last week and only an exchange of notes is now needed to activate the revised agreement.

The speedy action by the State Department reflects the pressures brought to bear by Japan. According to Alan Kuperman of the Washington-based Nuclear

Control Institute, an independent policy study body, the Japanese government was concerned that "a Dukakis administration might not be so compliant" on the issue of plutonium transport by sea and looked for a change in the agreement before the presidential election.

The task of shipping plutonium by air was made more difficult for Japan when Congress specified that casks containing plutonium must be able to withstand a worst-case air accident. An amendment added to a budget bill by Senator Frank Murkowski (Republican, Alaska) succeeded in further recommending that a jumbo jet laden with sample containers should be test crashed.

According to a State Department official the risks of sea transport are "greatly exaggerated" and safety "can easily be assured". Hisashi Dobashi of the Nuclear Safety Bureau of Japan's Science and Technology Agency says the question of whether ships or planes will be used is still "under study". The government is considering building a large patrol boat equipped with a helicopter pad to escort plutonium-carrying ships. But, in the long term, the government's sights are still set on air transport, which is much cheaper. This year's annual report on safety of nuclear energy released by the Nuclear Safety Commission last week concludes that air transport is a "strong candidate" and says that further research and development of crash-proof containers will continue.

Congress has not completely abandoned opposition to sea transport. Both the Foreign Affairs Committee and the Foreign Relations Committee have sent strongly worded notes to President Reagan asking for a new assessment of the risks and raising the possibility that Congress may begin new legislation to "address inadequacies in the treaty".

Alun Anderson and David Swinbanks

Battle of the research councils shows no sign of abating

London

BICKERING over the control of UK government-backed biotechnology broke out in public last week, leaving no doubt that effective collaboration between the Medical Research Council (MRC) and Science and Engineering Research Council (SERC) is as remote as ever, and that MRC and the Department of Trade and Industry (DTI) are also at loggerheads over biotechnology.

DTI's chief engineer and scientist, Ronald Coleman, began a two-day discussion meeting at the Royal Society by saying that positions were hardening and rivalry intensifying between the research councils. In a barely veiled attack on MRC, he criticized the inflexibility of permanently staffed research institutes devoted to curiosity-led research. In response, MRC secretary Dai Rees reminded his attacker that it was on innovatory research in MRC institutes that much of biotechnology was built, and that overmanagement stifles innovation.

Many speakers measured achievements in biotechnology against the hopes of the 'Spinks Report' of 1980, which recommended the measures needed to boost British industrial development in biotechnology. The report recognized that all the research councils had a legitimate interest in biotechnology and called urgently for a coordinated programme of research, and a greatly increased budget.

Reviewing subsequent events, Margaret Sharp of the University of Sussex's Science Policy Research Unit reminded the audience that the government of the day flatly refused to finance a gap that it felt industry should fill, and that MRC in particular blocked an effective coordination between research councils. In large part, she said, this was because Rees's predecessor, Sir James Gowans, was unwilling to push MRC in the direction of strategic research that SERC had adopted with the founding of their Biotechnology Directorate in 1981. Instead he strongly backed the foundation of Celltech, now Britain's leading biotechnology company, to exploit ideas emerging from MRC research.

"There are still problems of difference in ethos; ours works and delivers to industry", said Rees, who recently founded the Collaborative Research Centre (see below) as another way to exploit MRC discoveries. While Margaret Sharp argued for belated cooperation, saying it was particularly absurd that the protein engineering programmes of all the research councils and the DTI were not carried out as a national programme, Tom Blundell of Birkbeck College defended pluralism. "We would lose", he said, from an overseeing directorate that dictated one mode of research, adding that there was no lack of collaboration among the researchers themselves.

Mill Hill centre promotes innovation

London

THE Medical Research Council (MRC) Collaborative Centre, formally opened this week, is the council's latest way of helping to plug the notorious UK pre-development gap. Companies interested in an MRC invention will pay for a 1-3-year project at the centre to explore the invention's commercial value.

At present, about a dozen projects are in operation, but there is room for several more. Nearly all the current ideas have been generated at the MRC National Institute for Medical Research, which is on the same site in Mill Hill, on the northern outskirts of London. Most of the industry involved so far is British but a few US companies already sponsor projects.

Work in progress at the centre ranges from biosensors to protein purification, although a focus on antiviral agents and antibodies has emerged in the two-year run up to the official opening of the centre.

Typically, the project is carried out by staff specifically employed on short-term contracts, with the inventor acting more

or less as a consultant. There are no consultancy fees but an attempt is under way to circumvent the rules that prevent payment, says the centre's business manager, William Matthews.

For the company, which pays the full costs of the project including overheads, the perceived attractions of the centre include regular progress reports and a guarantee of full confidentiality, which is often not forthcoming from academic collaborators.

This year, MRC will still provide some 40 per cent of the costs of the centre, but it is on course for being self-financing by 1991. The centre will not generate profits. Instead, MRC should ultimately benefit by increased licence and royalty fees.

Making the most of the fact that Kenneth Baker, Secretary of State for Education and Science, was present to perform the opening ceremony, MRC's chairman Lord Jellicoe said that the centre looked forward to the day when it is receiving more alpha-rated projects than it can handle — a position already unfortunately reached by MRC itself. **Peter Newmark**

At the discussion meeting Sharp said that, contrary to some expectations, the strong industrial presence on the directorate's management committee "far from pushing towards the applied end (of biotechnology research) has done more or less the opposite". And with DTI now in general switching its support from 'near-market' to 'pre-competitive' research and development, firmer links than ever are being forged between the Biotechnology Directorate and DTI's Biotechnology Unit.

MRC may see this shift as a point in favour of its gaining control over SERC's biotechnology funds, which it has been eyeing for some time. But Sharp argued strongly for building on what was already in existence rather than switching horses in mid-course.

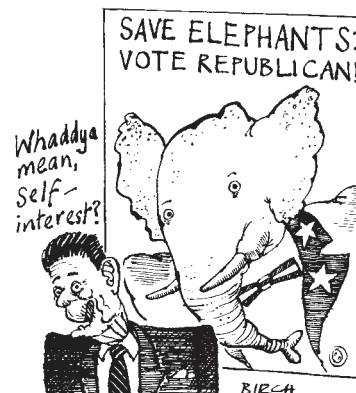
Peter Newmark

Endangered species

Washington

PRESIDENT Ronald Reagan has signed a bill to strengthen the 15-year-old Endangered Species Act. Besides toughening penalties for breaking the law, the bill lays out for the first time a scheme to protect the African elephant population, reported to have declined by nearly 50 per cent since 1979.

The bill gives the Department of Commerce \$1.5 million until the end of 1989 to



spot species at risk and develop a recovery plan for them. The Secretary of Agriculture will provide financial assistance to any state cooperating in the plan. And there is special emphasis on a follow-up programme to monitor for at least five years those species which have recently been recovered and removed from the endangered list. The bill raises the lowest penalty for violating the provisions of the act from \$10,000 to \$25,000, and the highest from \$20,000 to \$50,000.

A separate section of the bill sets up limits on imports of raw and worked ivory into the United States, to prevent the poaching of African elephants (see *Nature* 333, 290; 1988). In order to trade with the United States, ivory-producing countries must follow strict rules set up by the ivory control system of the Convention on the International Trade in Endangered Species.

Elizabeth Berendt

Nuclear start-up procedures revised in wake of failures

Washington

THE US Department of Energy (DoE) announced last week that it would try an entirely new approach to restarting its defence production reactors at the Savannah River Plant in South Carolina. The decision comes in the wake of revelations of continued safety problems at Savannah River and other DoE plants involved in the production of nuclear material for weapons. DoE Secretary John Herrington says the problems are coming to light because of heightened activity by environmental safety and health officials within the department, but critics say DoE is only reacting to a crisis situation, and there is no guarantee that the department's current interest in safety will translate to a new attitude within the department.

DoE's safety concerns have focused most recently on the problems associated with restarting one of the three heavy water reactors at Savannah River used to produce tritium. The Savannah River reactors are each more than thirty years old, and are showing signs of age. The plant has been operated by E. I. du Pont de Nemours and Company since it was built. But Du Pont has declined to renew its contract, and its role will be taken over by Westinghouse next April.

The P reactor was shut down last April to evaluate and upgrade the seismic capability of safety systems. Since the shutdown was expected to be brief, tritium that was being produced was left in the reactor. But the restart was delayed until 7 August and when operators attempted to restart the reactor they failed to take into account the decay of tritium to helium-3, a neutron absorber that can "poison" the reactor.

Operators could not understand why the control rods had to be removed so far in order to achieve criticality, and subsequent analysis showed that there was a potentially dangerous amount of reactivity that could not be accounted for. What upset DoE was that the operators did not stop the start-up procedure to find out what was wrong.

A build-up of xenon led to the reactor being shut down on 9 August, but it was restarted again that same night. On 10 August, an unexplained 2 per cent jump in power occurred, but operators decided to proceed with the start-up procedure. The reactor was finally shut down on 17 August.

Although DoE does not believe that the incident represented a danger to the public, the department was nonetheless chastened by these events. A memo written by Richard Starostecki, deputy assistant secretary for safety, health and quality

assurance, criticized the attitude of the plant operators as being "a prelude to disaster". He wrote that "the operating staff and advisors, knowing that they did not understand what was going on inside failed to put the reactor in a safe, known condition (shutdown) while they investigated, did not vigorously investigate, and THEY DIDN'T CARE!" (emphasis in memo).

Speaking at a press conference last week, DoE deputy secretary Joseph Salgado described the department as "in transition". He says DoE is moving into a period where safety will receive renewed emphasis, adding that the department is watching its contractors much more closely.

It will be some time before new safety systems can be installed in the P reactor, so DoE is now planning to start the K reactor by the end of the year. The K reactor has recently had a fourth emergency core cooling system installed, and is now undergoing seismic tests. For the first time, a phased start-up procedure will be used. Specific targets in the start-up process will have to be met to the satisfaction of DoE before the process will be permitted to continue. If all goes well, DoE hopes to have the reactor operating

by January of next year.

The United States currently has no plant capable of producing tritium for its nuclear weapons arsenal, but DoE officials are confident that one of the Savannah River reactors can be brought back on line before the situation becomes critical. DoE has denied that it may be necessary to 'cannibalize' weapons to obtain tritium.

Although DoE has been critical of Du Pont's operation of Savannah River, there are those who feel that DoE must ultimately be held accountable for the problems at its facilities. At a hearing at the end of last month Representative Mike Synar (Democrat, Oklahoma), chairman of the environment, energy and natural resources subcommittee of the Committee on Government Operations spoke of an "attitude of indifference to safety and an unwillingness by DoE, even at the highest levels, to assume control of its own facility". For its part, Du Pont chairman Richard Heckert has rejected charges that his company has operated the plants in an unsafe manner, pointing to a 38-year safety record.

Savannah River is not the only DoE weapons facility with safety problems. DoE officials last week announced the shut-down of a key portion of its Rocky Flats materials processing plant in Colorado because of poor safety procedures, and another fuel processing facility at Fernald, Ohio, is currently in the midst of a labour dispute.

Joseph Palca

Peaceful approach to Northwest Passage

Washington

IN the first test of an Arctic agreement signed by the United States and Canada in January, the US administration has asked for permission for a US vessel to pass through the disputed Northwest Passage. The agreement is a compromise that dictates the United States must ask to enter Arctic waters claimed by Canada even though it refuses to acknowledge Canadian sovereignty over them.

The vessel involved is the US Coast Guard icebreaker *Polar Star*, which is retreating into Canadian-claimed waters after failing to break through unexpectedly heavy ice off the coast of Alaska. The vessel is the sister ship of the *Polar Sea* whose unannounced transit through the Northwest Passage in 1985 outraged Canadians and triggered an international dispute (see *Nature* 333, 733; 1988).

The United States, backed by NATO countries, wished to regard the Northwest Passage as international waters, partly because of their military importance for the movement of submarines between the Atlantic and Pacific Oceans and partly because granting sovereignty there would lead to similar claims on waterways elsewhere in the world. The Soviet



The Northwest Passage winds among Canada's islands from the Prince of Wales Strait to Lancaster Sound.

Union supported Canada's view as it claims similar territorial rights in its own Arctic territories.

The dispute made Canada aware that it had no way to police the Arctic territories and led to plans for a massive all-year capable ice breaker and for a fleet of British or French nuclear submarines capable of penetrating beneath the ice.

Alun Anderson

Polish government allows tentative steps towards academic autonomy

London

"THERE can be no real development of science or our country without academic self-governance", Polish politbureau member Marian Orzechowski stated last month. His remark launched a series of concessions on academic autonomy that go a long way towards reversing the clamp-down of the 1985 Higher Education Act. In particular, although the pro-Solidarity Independent Students' Association (NZS) is still outlawed, for the first time since the imposition of martial law in December 1980, some student organizations not affiliated to the official Association of Polish Students (ZSP) will be allowed to operate at the inter-university level.

Orzechowski was addressing the Main Council for Science and Higher Education, a 70-member elected body representing all Polish universities and higher colleges (excluding the Catholic University of Lublin which opted for observer status only). The Main Council was established during the Solidarity era (1980-81) and under the 1982 Higher Education Act the government was obliged to obtain its consent to major decisions on higher education policy. Under the 1985 Act, however, it was demoted to a consultative role only. Last month's meeting, while not restoring the Main Council's teeth, converted it into a more active forum for the

exchange of ideas.

The Polish government is at present considering a number of changes in the law on higher education, which as Education Minister Henryk Bednarski told the Main Council, include the establishment of a Fundamental Research Council, described as a forum in which the voice of every scientist would carry the same weight, and new arrangements for the funding of research in the universities. The most pressing problem, however, is to restore an atmosphere of confidence within the universities.

Poland's endemic economic crisis has depressed both academic salaries and student stipends, and libraries and laboratories still depend on the generosity of Western colleagues for essential journals and equipment. Many talented young people are now reluctant to apply to university since by so doing they will merely commit themselves to five years of poverty in poor hostel conditions, with little or no hope of a decent salary at the end.

Moreover, the students have lost confidence in their official union, the ZSP. Less than 10 per cent of students belong to the official union, even though it tries to restrict certain material benefits to members only. During the past few years, students have called repeatedly for the legalization of the banned NZS, and in some universities, NZS now operates openly, running bookstalls and holding meetings. Warsaw University's rector, Dr Grzegorz Bialkowski, has declared that he, for one, is prepared to liaise with NZS even now, on the grounds that it is the only genuine representative student opinion in his university.

At the end of September, a national conference of the ZSP admitted the need for "pluralism" of student organizations, not so much a concession as a pragmatic admission of its own weakness. A few days later, the first independent student organization since the Solidarity era was officially registered in Poznan, the Catholic-orientated Academic Union of Young Poland (ZAMP). A few days later, a similar organization, called Verbum, was founded in Gdansk. Furthermore, at the celebrations marking the opening of the academic year in Poznan last week, attended by party leader General Wojciech Jaruzelski, the rector of the Adam Mickiewicz University of Poznan, Dr Jacek Fisiak, (since appointed minister of education) urged his students to take advantage of the new pluralism, which, he said, permits any student organization with "productive aims" providing it is not opposed to the constitution. **Vera Rich**

Shroud a good forgery

Zürich

AFTER weeks of rumour and speculation, the official carbon dating results for the Turin Shroud were released in Zürich last week. They show that the shroud, claimed by its proponents to be the burial shroud of Jesus Christ, is at most 728 years old, originating "with 95 per cent certainty" between AD1260 and 1380. Three laboratories at the Swiss Federal Polytechnic Institute, at the University of Arizona, and at Oxford University arrived at the dates in independent analyses.

The Catholic Church, which for years had refused to allow the shroud to be tested using destructive means such as radiocarbon dating, accepted the researchers' finding without dispute. The Archbishop of Turin, Anastasio Cardinal



Ballestrero, confirmed the symbolic value of the shroud despite its lack of authenticity. The dating of the shroud was performed on just a few square centimetres of cloth taken from the edge of the linen shroud in April of this year, making use of techniques perfected during the past decade which rely on single atom counting using a particle accelerator.

Like traditional radiocarbon dating methods, the technique relies upon determining the ratio of carbon-14 to carbon-12 isotopes in the sample. Living organisms accumulate the rarer carbon-14 isotopes during their lifetimes, but these unstable atoms decay over time (with a half-life of about 5,500 years) and cease to be replaced once an organism dies. The current method can determine the ratio to better than 1 per cent accuracy, corresponding to an uncertainty of less than 80 years either way.

Though the new results seem likely to end centuries-old speculation about the origins of the Shroud (a 1389 letter from the Bishop of Troyes to Pope Clement VII claimed it was a forgery), the source of the Christ-like image remains a mystery.

Steven Dickman

Leicester makes bid for space eminence

London

THE British 'space lobby' is continuing its efforts to strengthen Britain's position in space despite the government's lack of enthusiasm. Leicester University this week launched a £3 million appeal for a new space centre. It will be self-financing, although initially funds will come from industrial and commercial investment. Due to be built and begin operation in 1990, the centre will host the country's first undergraduate degree course in space science and technology. It will be built on the same foundations as traditional physics courses but "will have a different flavour" says Professor Kenneth Pounds, head of physics. Courses will include space-craft systems design, space instrumentation, earth-observation science and orbital dynamics and space propulsion.



Flavour enhancer.

Christine McGourty

The debate continues

SIR—In the test system used by Benveniste and his co-workers¹ there was another source of anti-IgE which the authors were apparently not aware of — namely the basophils themselves. In recent years naturally occurring auto-anti-IgE antibodies of the IgM^{2,3} and IgG⁴⁻⁸ classes have been described in a large proportion of normal people and patients with allergic manifestations. Data on the biological and physiological effects of auto-anti-IgE in man and rodents are now being reported at an increasing rate and are contained in a current volume devoted to this subject⁹. IgG anti-IgE, unlike IgM anti-IgE autoantibody, is directed against the heat labile D₂ antigen of human Fc. It circulates as a free molecule or in immune complex form with self IgE. IgE-IgG anti-IgE complexes are found on mast cells and basophils where they are anchored to the membrane via the Fc portion of IgE. Amongst other *in vivo* and *in vitro* activities, IgG anti-IgE autoantibody can degranulate basophils and mast cells much as heterologous anti-IgE does.

Benveniste and his team did not exclude spontaneous basophil degranulation by auto-anti-IgE, a possibility supported by their admission to *Nature*'s team that the best results were obtained when cells were left overnight before counting. The fact that not all people have high or detectable IgG anti-IgE autoantibody explains why some blood samples gave negative degranulation results. Lack of degranulation in the control, where anti-IgG was used instead, does not invalidate this explanation as the control was carried out at a different time from the test run. Besides, there have been periods when cells did not regranulate even with anti-IgE.

Benveniste and co-workers should now repeat their experiments using cells stripped of their IgE and re-sensitized with known uncomplexed IgE.

FAROUK SHAKIB

*Asthma and Allergy Treatment and Research Centre,
12 Vernon Street,
Derby DE1 1FT, UK*

1. *Nature* **333**, 816–818 (1988).
2. Williams, R.C. *et al.* *J. clin. Invest.* **51**, 955–963 (1972).
3. Magnusson C.G.M. & Vaerman J.P. *Int. Archs. Allergy appl. Immunol.* **79**, 149–156 (1986).
4. Inganäs M., Johansson S.G.O., Bennich H. *Int. Archs. Allergy appl. Immunol.* **65**, 51–61 (1981).
5. Nawata, Y. *et al.* *Clin. exp. Immunol.* **58**, 348–356 (1984).
6. Nawata, Y. *et al.* *J. Immunol.* **135**, 478–482 (1985).
7. Quinti, I. *et al.* *J. Allergy Clin. Immunol.* **77**, 586–594 (1986).
8. Wilson, P.B., Fairfield J.E., Beech N. *Int. Archs. Allergy appl. Immunol.* **84**, 198–204 (1987).
9. Shakib, F. (ed). *Monogr. Allergy* **26** (Karger, Basel, in the press)

SIR—The editorial staff of *Experientia* has followed with great interest and fullest sympathy the 'alarums and excursions' following your publication of the by now famous (or infamous?) Benveniste report.

We found your singular and 'unorthodox' decision to include James Randi in the investigation team courageous. Recently (*Experientia* **44**, No. 4, April 1988) we published a multi-author review, coordinated by Professor David Marks, on 'Investigating the Paranormal'. One of the authors was Randi, and he described in vivid language and detail the 'debunking' of 'psychic' events at a widely publicized 'television preacher's' meeting. From what we know of Randi, his almost uncanny ability to detect fraud (whether conscious or unwitting) from his great experience in practical conjuring makes him an ideal investigator, quite apart from his intellectual honesty.

Surely it is the duty of multidisciplinary journals such as *Nature* and *Experientia* not only to publish 'serious', 'first-rate' science, but to provide a forum for imaginative — and thus risky — ventures into new fields, stimulating others to investigate the phenomena described.

H. P. VON HAHN

Experientia, PO Box 133,
CH-4010 Basel, Switzerland

SIR—If we are to continue to hear arguments about the Benveniste findings, perhaps a *reductio ad absurdum* argument should be considered. One needs only to drink high purity deionized water to have a cure for all diseases subject to chemical therapy, as all chemicals are present in this pure water in infinite dilution.

JACK P. SHOUP

*4705 Brady Place,
Colorado Springs, Colorado 80915, USA*

SIR—The stage magician's first rule is to divert the audience's attention from what is actually taking place. It therefore needs to be mentioned that the presence of a skilled stage magician during a scientific test is a double-edged sword. The presence of James Randi during *Nature*'s observations of Benveniste's experiments would seem to guard against any sleight-of-hand on the part of Benveniste's staff. Unfortunately, it is also the case that Randi could easily have used his skill to influence or even determine the apparent outcome of the experiments, with nobody present realizing it.

This is not a moot point, as Randi makes no secret of the fact that he has interfered intentionally in experiments involving 'fringe areas' of science. For example, a widely reported New York news conference called by Randi in January 1983 disclosed his use of two young magician associates to subvert the experiments of Dr Peter Phillips, a physicist at the McDonnell Laboratory for Psychical Research in St Louis, Missouri.

How are we to be any less sceptical of

the *Nature* review, in which Randi played a prominent part, than we were of the original Benveniste report?

DAVID DUNTHORN

*908 West Outer Drive,
Oak Ridge, Tennessee 37830, USA*

SIR—Another disturbing aspect of the Benveniste affair is your statement that the decision to publish the original paper was based partly on its coming from the head of a major laboratory and that similar papers from unknowns would be dealt with less positively. To the extent that this attitude prevails among editors (and I am one) it must contribute to the suppression of genuinely unconventional advances.

LEIGH M. VAN VALEN

*Department of Ecology and Evolution,
University of Chicago,
Chicago, Illinois 60637, USA*

Availability of theses

SIR—The correspondence started by Beverly Halstead (*Nature* **331**, 497; 1988) about the training of PhD students and the value of theses has raised many fundamental issues, such as whether theses should consist of previously published articles or not. One of the arguments presented in favour of publication was the inaccessibility of PhD theses in their present form.

The British Library Document Supply Centre at Boston Spa has for the past 18 years been trying to improve the availability of British doctoral theses. Almost all UK universities now send their theses to the British Library for microfilming. These films or enlargements thereof are then available for loan or retention.

Information on the 75,000 theses now held and on the 5,000 being added to stock each year is contained in the library's monthly publication *British Reports, Translations and Theses*. Since 1985, bibliographic details of new theses have also been added to the SIGLE (System for Information on Grey Literature in Europe) online database accessible through BLAISE. From this year, details of theses held at Boston Spa will also be included in the Dissertation Abstracts International databases produced by University Microfilms International.

Despite some assertions (see *New Scientist*, 21 January) that few theses are consulted by more than four people (author, supervisor and two external examiners) the Document Supply Centre provided copies of more than 13,000 during 1987.

D.N. WOOD

GARTH FRANKLAND

*British Library,
Document Supply Centre,
Boston Spa, Wetherby,
West Yorkshire LS23 7BQ, UK*

Is the literature about to be readable?

Dissatisfaction with the standard of today's scientific literature is rife. It does not follow that change will happen, but only that it might.

If the scientific literature is indeed the chief way of setting and maintaining professional standards in science, there is at least a chance that beneficial changes are on the way. The article by J. D. Watson and F. H. C. Crick on the structure of DNA, which appeared in *Nature* in 1953, would probably not now be publishable. That article occupied just over half a printed page. While there are references to other work on the storage of genetic information, the suggestion that it must be DNA emerges like a rabbit from a hat. There is no substantial discussion of the implications of the proposal, merely the now-famous sentence: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material" (*Nature* 171, 737–738; 1953). Today's referees would have clamoured for a fuller account of the reasons for considering the proposed structure seriously and an account of the implications.

The reasons why the conventions of the scientific literature have so radically changed are well-known, as are the consequences. Most scientific communications have the same ingredients. Authors describe their problem and discuss its antecedents, give the rationale for their own attack on the problem and the techniques which they have used and finally discuss the significance of the results. There are, of course, many minor variations on the theme; experimental details may be printed separately from the rest of the text, or appear as figure legends. But the general architecture of a paper is constant.

Or, at least, it has been. There are now signs of discontent with the literature as it is. The sheer bulk of it is not what matters most; it matters more that the literature may not be suited to its purpose. The prospect for change may now be brighter than for decades.

While electronics will have an important influence on the speed and efficiency of publication, electronics and computer technology are not the central issues. For while there are certain to be some fully electronic journals which can be read only by sitting at a keyboard, most of the literature is likely to remain for a long time as ink on paper. People like to see what they have done; an accession number from a data-bank is not a sufficient substitute.

Meanwhile, it is likely that the changes (if any) that come about will consist of a

reordering of the priority at present attached to the conflicting functions of the literature, which are several. Thus the published literature is a kind of record of the accumulation of discovery from which, for example, historians may reconstruct the development of innovation. The literature is also a means by which those who would make use of science can gain access to original research, although most technology transfer of this kind entails the use of specialized magazines or trade journals. But there are strictly peripheral functions. The published literature is the means by which the research profession's scepticism is brought to bear on new developments, helping to tell the wheat from the chaff; the claim is undoubtedly correct, but would be more easily accepted if the process were more explicit.

None of this explains why the format of the standard scientific article is what has now become conventional. Historians, for example, would prefer detail of a different kind, those concerned with the application of science would prefer that the literature should be more intelligible, and so on. But the more substantial reason why the literature is in its present form is that the research profession, like other scholarly professions, but more so, has come to rely on the published literature for the continued setting and maintenance of professional standards. Bibliographies are a prominent part of job and promotion applications, while a published 'track record' helps in telling which applications for research grants will succeed.

Researcher's assessments of each others' qualities rest, in the last resort, on the published record. Naturally, there are other ways in which the same job might be done; engineers, physicians and the like have professional associations for setting and maintaining standards, for example. In science, where surprising innovation from unexpected sources is welcomed, the use of the published literature for this purpose makes more sense, but also gives it its distinctive characteristics and makes it hard to read.

This is why referees so often give authors a hard time and usually do so gladly, even though theirs is the most thankless of all tasks. In a profession which is necessarily elitist, paying close attention to the quality of what others do is plainly a professional task transcending particular questions of which articles

should be published in which journals. This is the spirit in which referees demand that references to earlier work should be fair (whatever that may mean) and that authors should be restrained in canvassing the implications of their work: unbridled speculation is a means of laying claim to all kinds of ideas not directly suggested by the data gathered.

Sadly, these are precisely the points on which the contradictions of the literature are most clamant. Speculation of the kind that referees dislike may nevertheless help enormously to define an author's long-term objectives or to explain why he set about his task in a particular way. Detailed accounts of experiments, necessary so that others can set about the repetition of published work (but rarely quite sufficient), are not often the details that most users of the literature would wish to have. And the literature would be written differently, and more palatably, if its function in setting standards was not as dominant as it has become.

The signs that changes may be under way are, admittedly, only straws in the wind. Some appointments committees in the United States have begun limiting the allowable size of a candidate's bibliography. There is also a growing uneasiness that the literature may sometimes be manipulated, especially in relation to the assessment of attainment, illustrated by the several cases of fraud brought to light in the United States during the past few years.

There are also more substantial considerations working for reform, one of which is the huge amount of data now unpublished in any formal sense, in fields as different as geophysics and the projects for collecting the nucleotide sequences of genes. Can the long-term interests of the process of discovery be properly served if the bulk of this data remains unanalysed and unpublished because of the convention that the Minimum Publishable Unit should include at least an attempt to make sense of some problem in science. The fact that, in other fields, preprints have replaced the published literature as means of communications is another sign of the pressure on the present system.

Luckily, these are all fields in which electronics may provide both solutions and new opportunities. Whether the literature will ever fully revert to its literary purposes is, of course, another matter.

John Maddox

Protein structure

Trapping folding intermediates

Christopher M. Dobson and Philip A. Evans

COULD a protein fold by wandering without direction through conformational space until, chancing upon the native structure, it drops precipitously into the potential well that lies there? After Cyrus Levinthal estimated just how long this might take¹, the idea came to seem implausible, and although reports of its death may have been greatly exaggerated^{2,3}, it is now widely held that there must be pathways of protein folding, leading to the native state via specific intermediates⁴. Unambiguous detection, let alone detailed structural characterization, of such species has turned out to be an exacting challenge, but one which is now being met through some ingenious experimental strategies. One such idea, originally suggested by Robert Baldwin

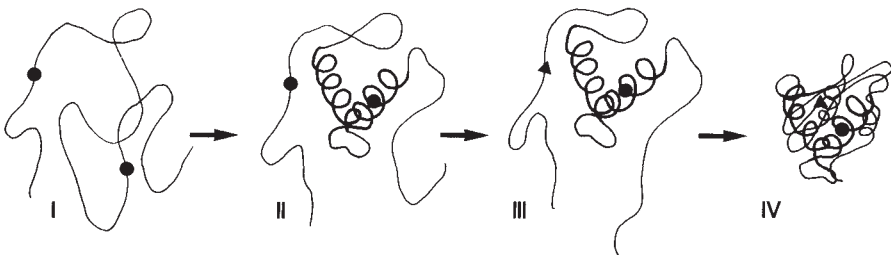
by NMR and by monitoring the absorbance of derivatives with selectively nitrated tyrosines^{8,9}. An alternative approach has been devised by Adler and Scheraga, who constructed a flow system which delivers a refolding protein into, say, an NMR spectrometer within a few seconds of initiating the reaction and then feeds it continuously back through the unfolding/refolding cycle, so permitting signal averaging¹⁰. Good-quality spectra of an intermediate in the refolding of ribonuclease A have been obtained in this way.

So far, however, the most detailed structural characterization of folding intermediates has been achieved by opting out of the race against the clock and developing instead means for trapping them, so that subsequent analysis can be

with the peptide model; this suggests that we do not really understand what is happening to the parts of the molecule not yet in their native-like conformation in the intermediate. The pancreatic trypsin-inhibitor system thus promises to continue to be a key system in the investigation of the folding problem.

The work of Udgaonkar and Baldwin and of Roder *et al.* described in this issue^{6,7} represents a variation on the trapping theme. The idea is to compare the stage in the folding reaction at which different exchangeable protons become protected from the solvent, presumably as a result of structure formation. The figure shows how this is achieved. The experiment was originally conceived as a straightforward competition between refolding and hydrogen exchange, as monitored by tritium labelling. Its evolution into a high-resolution probe of folding intermediates has depended on the development of a pulse-labelling methodology and on advances in NMR spectroscopy. It is now possible to resolve and assign the

Schematic view of the experiment used to characterize folding intermediates. I, The protein is initially unfolded in a concentrated guanidinium chloride solution in D₂O. All exchangeable protons become substituted by deuterium under these conditions; two representative examples are indicated by black circles. II, Refolding is initiated by dilution of the solution. Early folding intermediates may be formed such that a proportion of exchangeable hydrogens becomes protected from the solvent — this is shown schematically for one of the two representative hydrogens. III, After a short interval, the labelling pulse is initiated by dilution into an excess of H₂O. The pH is kept high at this point, favouring base-catalysed hydrogen exchange and leading to rapid protonation of all amides (indicated by black triangle) except those protected by the structure of the folding intermediate. IV, The labelling pulse is terminated by lowering the pH close to that where hydrogen exchange rates are slowest, so that refolding can now proceed to completion without significant further exchange. Ultimately, many exchangeable hydrogens will be trapped in the folded structure. For these, the extent of labelling by protons can be determined at leisure by ¹H NMR spectroscopy. In this case, the site which was protected in the early intermediate would have little or no proton incorporation (circles), while the other, which was structurally excluded only later in folding, will have significant proton intensity in the spectrum (triangles). By varying the delay before introducing the label, the sequence of structure formation can be investigated, as shown, for example, in the two papers in this issue.



and colleagues⁵, has now come to fruition as described in two papers in this issue, one, on page 694, from Baldwin's own group at Stanford University⁶ and the other, on page 700, from Heinrich Roder and colleagues at the University of Pennsylvania⁷.

The problem with kinetic folding intermediates is their ephemeral nature. Folding of small proteins often occurs in less than one second and, even if it is slowed down by a need for proline isomerization or disulphide crosslinking, time is generally too short for conventional application of powerful but slow structural techniques like nuclear magnetic resonance (NMR) spectroscopy. In a few cases it has been possible, nevertheless, to carry out direct spectroscopic studies of transient intermediates. Biringer and Fink have, for example, slowed right down the refolding of ribonuclease A by carrying out the reaction at -30°C in aqueous methanol solution. Under these conditions, intermediates accumulate which the authors studied

carried out at leisure. The classic case of this is Tom Creighton's elucidation of the folding pathway of the bovine pancreatic trypsin inhibitor, as reflected in the sequence of disulphide formation in the course of oxidative refolding¹¹. His results not only revealed that there are specific intermediates but permitted the trapping, isolation and structural characterization of a series of them, on the basis of which a picture of molecular events in the folding pathway has been put forward¹².

Some of these structural conclusions have now been tested by Peter Kim and Terence Oas, who show that two peptide fragments from this protein form a complex with native-like structure when they are crosslinked by a disulphide bridge¹³. This seems to be analogous to what happens in the formation of a key intermediate, having this same disulphide, in the protein-folding pathway. But there are interesting differences in, for example, the apparent cooperativity of the folding interactions of the partially folded protein compared

resonances of many amide protons through two-dimensional NMR and to use their intensities as a measure of the extent of substitution by deuterium which has occurred during a labelling pulse.

The authors of the papers in this issue concur in the general conclusions to be drawn from their independent studies of two small proteins: the time-dependences of the acquisition of protection from solvent exchange are distinct for different amides and are multiphasic. The former observation clearly excludes the single co-operative folding event as a viable model; structural intermediates must be formed in which some but not all amides become protected before formation of the final tertiary structure. The multiphasic kinetics are less readily explained, but may reflect heterogeneity of the unfolded state or the existence of a small number of alternative folding pathways. More work is needed to resolve this question. In cytochrome *c*, amides from two specific helices have been pinpointed as becoming protected

in an early folding event⁷; in ribonuclease A, amides of the β -sheet are among those which acquire early protection⁶. The authors of both papers argue that these results are consistent with the 'framework' model⁴, which postulates that elements of secondary structure are formed early in folding, with subsequent packing of these pre-formed units to form the native structure. It is important always to bear in mind, however, that the secondary structure could be supported by quite extensive side-chain interactions in these early intermediates, though these are not detected directly by this experiment. As Roder *et al.* point out⁷, it is noteworthy in this context that the first stabilized helices of cytochrome *c* interact closely in the structure, and the stability of the intermediate could rest as much on the mainly hydrophobic inter-helix interactions as on the intrinsic tendency of the individual chain segments to fold in this way.

There is a touch of irony about the role of hydrogen exchange in these studies. Some of the earliest evidence for the presently accepted dynamic view of protein structure came from the observation of the exchange of nominally inaccessible hydrogens with the solvent, and interest has long focused on the kinetics of these events. In the present work, the interest is in how quickly these hydrogens can be stopped from exchanging. However, the point that exchange can still happen even when a hydrogen appears to be structurally protected has not been lost — Udgaonkar and Baldwin show⁶ for several amides that the measured extent of protection is independent of the duration or intensity of the labelling pulse, so that once it has formed, the protective structure is exceedingly stable.

This result is really rather remarkable in view of the conventional idea that thermodynamic and kinetic stability rests on the context of a structural element within the full native conformation. At the other extreme, we should bear in mind that some species with a great deal of secondary structure nonetheless afford only marginal protection from solvent exchange — this is true of the molten globule state of α -lactalbumin, which has extensive helical structure and a condensed

hydrophobic core, lacking only a defined tertiary structure, but which has only a small group of significantly protected amides¹⁴. It is not completely clear what is needed to provide effective protection from exchange — hydrogen bonding may be neither required nor sufficient. That being so, it might be an oversimplification to assume that those parts of a protein which are not highly protected early in folding have simply not folded.

The cooperativity of folding has long seemed to be a formidable obstacle in the way of attempts to dissect the architecture of globular proteins. The outlook is brightening markedly, however, both

with these successes^{6,7} in teasing out some of the details of the kinetic pathways of refolding, and with the realization that species such as structured peptide fragments and partially organized states of proteins do exist and can profitably be studied at equilibrium. All this augurs well for the understanding of natural protein structures and the prospects for designing new ones. □

Christopher M. Dobson is in the Inorganic Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX1 3QR, and Philip A. Evans is in the Department of Biochemistry, Cambridge University, Tennis Court Road, Cambridge CB2 1QW, UK.

Planetary systems

Astrophysical effects of planets

Charles Alcock

THE expectation that planetary systems such as our own are commonly present around other solar-type stars is a reasonable speculation. Efforts to detect the presence of planets as massive as Jupiter around a few nearby stars (usually by examining the motion of the star) have thus far been unsuccessful, although the techniques are improving rapidly and either detections or interesting non-detections may be reported soon. Some theoretical work suggests that 'solar systems' with a few massive planets typically form during the birth of a star. In a recent paper¹, Struck-Marcell examines the role that such a solar system could play during the very late stages in the evolution of the star. After nuclear burning exhausts the hydrogen in the core of the star, it swells rapidly to become very luminous (10^3 – 10^4 times more so than the Sun), but with a much reduced surface temperature (typically about 2,300 K). When this happens to our Sun, the Earth will be destroyed as it is engulfed in the hot atmosphere of the expanding star. The giant planets, most prominently Jupiter, may survive this epoch. The effect of such survivors in other solar systems may already be visible.

Most red giant stars lose mass at copious rates, in some cases exceeding 10^{-6} solar masses per year ($M_{\odot} \text{ yr}^{-1}$): the more extremely evolved supergiant stars sometimes lose about $10^{-4} M_{\odot} \text{ yr}^{-1}$. In this outflowing material, a significant fraction of the stellar silicon is in the form of silicon monoxide (SiO) radicals. Radioastronomers have detected microwave emissions produced by many of the rotational transitions of the SiO molecule; the emitted power is so intense that the emission lines must result from the maser (stimulated emission) process, whereby enormous amplification of microwave radiation occurs. Thus far there is no good theoretical explanation of the SiO maser process

in these objects, although it has been possible to conclude that a huge amount of SiO must be present^{2,3}.

Struck-Marcell does not discuss the excitation of the maser itself, but the model he proposes does concern the general morphology of these sources of radiation. Specifically, there seem to be only a few sources of SiO maser radiation associated with each red giant star and these seem to be spread over a region $\leq 10^{14}$ cm in radius (our Solar System has a radius of about 6×10^{14} cm). The radial velocities of the various maser regions are spread over 10–20 km s⁻¹. Many of the sources are highly polarized. Struck-Marcell proposes that planetary systems (similar to ours) are present in each of these red giants and supply the SiO molecules that are observed as masers. This hypothesis is unique in the maser literature and highly unusual in the whole of astrophysics: Struck-Marcell points out that planetary systems could plausibly have dramatic consequences for what is observed around stars, whereas conventionally they are expected to have only minor consequences.

The principal features of Struck-Marcell's model are as follows. A planet modifies the gas outflow from the central star by creating a supersonic wake — this certainly happens around the planets in our Solar System — and by the direct addition of material rich in silicon and oxygen, a process similar to the eruption of material from Jupiter's satellite Io, but in this case driven by thermal evaporation. As the SiO density most sensitively affects the microwave emission, the latter is probably more important than the former.

Struck-Marcell notes that with a conventional model for a red giant star, there would be insufficient heating of solid bodies for this process to work at radii of around 10^{14} cm. But he cites the recent

1. Levinthal, C. J. *chem. Phys.* **65**, 44–45 (1968).
2. Harrison, S. C. & Durbin, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4028–4030 (1985).
3. Dill, K. A. *Biochemistry* **24**, 1501–1509 (1985).
4. Kim, P. S. & Baldwin, R. L. A. *Rev. Biochem.* **51**, 459–489 (1982).
5. Schmid, F. X. & Baldwin, R. L. *J. molec. Biol.* **135**, 199–215 (1979).
6. Udgaonkar, J. B. & Baldwin, R. L. *Nature* **335**, 694–699 (1988).
7. Roder, H., Elöve, G. A. & Englander, S. W. *Nature* **335**, 700–704 (1988).
8. Biringer, R. G. & Fink, A. L. *Biochemistry* **21**, 4748 (1982).
9. Biringer, R. G. & Fink, A. L. *Biochemistry* **27**, 301 (1988).
10. Adler, M. & Scheraga, H. A. *Biochemistry* **27**, 2471 (1988).
11. Creighton, T. E. *Prog. biophys. Molec. Biol.* **33**, 231–297 (1978).
12. States, D. J. *et al. J. molec. Biol.* **195**, 731–739 (1987).
13. Oas, T. G. & Kim, P. S. *Nature* (in the press).
14. Dolgikh, D. A. *et al. Eur. Biophys. J.* **13**, 109–121 (1985).



100 years ago

THE following notices of anthelia may be interesting to the readers of NATURE. Frances Kidley Havergal thus described a sunset on the Faulhorn: "At one juncture a cloud stood still, apparently about two hundred yards off, and we each saw our own shadow gigantically reflected on it, surrounded by a complete rainbow arch, a full circle of bright prismatic colours, a transfiguration of our own shadows almost startling; each, moreover, seeing only their own glorification".

S. T. Coleridge described the phenomenon: "Such thou art, as when The woodman winding westward up the glen At wintry dawn, where o'er the sheep-track's maze

The viewless snow-mist weaves a glist'ning haze, Sees full before him, gliding without tread, An image with a glory round its head: The enamoured rustic worships its fair hues, Nor knows he makes the shadow he pursues."

Benvenuto Cellini saw, probably, this phenomenon, and supposed it peculiar to himself. F. Robertson cites it as a proof of inordinate vanity. He says: "Conceive a man gravely telling you that a vision of glory encircled his head through life, visible on his shadow, especially on the dewy grass at morning, and which he possessed the power of showing to a chosen few."

From *Nature* 38, 619; 25 October 1888.

work of Bowen⁴ who shows that a pulsating red giant has a much more extended atmosphere than a non-pulsating star, and that all of the red giants with SiO masers are pulsating. This extended, hot atmosphere drives the evaporation.

Having concluded that direct addition of SiO is the most promising source of the maser-emitting material, Struck-Marcell suggests that 'gallilean satellites' are the sources. Although it is certainly true that such objects are plausible sources for his model, any substantial rocky object, such as a rich, well-placed belt of asteroids or an appropriately located terrestrial planet, would do equally well.

The model does make some predictions. There should be only a few isolated maser-emitting volumes (corresponding to a few planets) around each star. Time variations should occur in the radial velocities of the emission lines as the host planets orbit their central stars. With the advent of very-long-baseline interferometry monitoring programmes⁵, spatial motions of the planets might be detectable. □

1. Struck-Marcell, C. *Astrophys. J.* **330**, 986-991 (1988).

2. Langer, S.H. & Ross, R.R. *Astrophys. J.* **284**, 751 (1984).

3. Alcock, C. & Ross, R.R. *Astrophys. J.* **310**, 835-841 (1986).

4. Bowen, G.H. *Astrophys. J.* **329**, 299-317 (1988).

5. Lane, A.P. *IAU Symp.* **110**, 329-333 (1984).

Charles Alcock is in the Institute of Geophysics and Planetary Physics, Lawrence Livermore National Laboratory, PO Box 808, Livermore, California 94550, USA.

Behavioural ecology

Tampering with territories

Robert M. May and Paul H. Harvey

WHAT criteria do males use for choosing a territory? The answer is determined partly by what females look for when they are choosing a territory (unless they decide purely by the characteristics of the male). A favourite question among behavioural ecologists has been to define territory quality, the characteristics of which are now known to differ among species¹. When males and females remain in a territory to raise the young, the territory is often largely determined by resource use. For example, large birds or mammals living on widely dispersed food items have the largest territories²⁻⁴. Among species that have no paternal care, males may set up mating territories at places where oestrous females are likely to pass through⁵. An extreme case is found among lekking species, where males collect together on hotspots, places where several female home ranges overlap⁶. Females from the different ranges come together at the lek and choose among the displaying males. It seems that, according to species, the position a male occupies at the lek may or may not influence his success at attracting females⁷. If the same sites are associated with high male reproductive success in successive seasons, irrespective of the males occupying them, it might seem natural for behavioural ecologists to seek the common characteristics of those sites. Warner's study of bluehead wrasse, reported on page 719 of this issue⁸, opens up the intriguing possibility that the main characteristic shared by successful sites can be that they have been

successful in the past.

Warner's experimental technique is both elegant and novel. He removed whole populations of fish from coral reefs and replaced them by roughly equally sized populations of fish from other reefs. It turns out that males are territorial at particular types of site, which are quite characteristic, but only a proportion of sites in an area is occupied by the local population. When populations are tracked through time, the same territories are occupied in successive seasons, sometimes by different males. When populations are transplanted, however, some new sites are used while some of the old ones are left vacant.

These transplant experiments open up the possibility of similar studies in other groups, particularly in lekking birds. They also suggest new ways of testing ideas about the spatial geometry of territories, and especially how this geometry can depend on whether territories are first established by the simultaneous arrival of individuals (at the start of a breeding season, for instance) or by sequential arrival.

John Maynard Smith has discussed models for different kinds of territorial behaviour and their implications⁹. In these simple models, there are two phases in the setting up of a territory. In the first phase, animals settle at random or at likely sites and start to defend territory boundaries. If another animal settles nearby, say to the north, while a third settles some distance away to the south, the original occupant



Territory boundaries of male mouthbrooder fish (reproduced from page 272 of ref. 11).

would move its northern and southern territory boundaries towards the south. As new animals arrive in the area, territory boundaries would change. After some time has elapsed, the animals enter a second phase, in which each individual selects a nest site in the centre of its territory: from this point on only territory boundaries can change. In either phase, if the territory is smaller than some predetermined size, animals abandon their territories.

One interesting point to emerge from Maynard Smith's model is a possible relationship between average territory size (and thence the number of animals to be accommodated) and the degree of synchrony with which territories are established. Clearly, if all individuals arrive together and go through the first phase before embarking on the second, the number of individuals finally settling will be the total area divided by the territory size; for example, when dealing with a one-dimensional habitat such as a stream or hedgerow (or the edge of a reef) of length L , and if the minimum territory has length $2r$, this density is $L/2r$. Alternatively, if the animals arrive in groups, and if each group passes on to the second phase before the arrival of the next group, then the final number of territories can be as few as one-half that generated by simultaneous arrival; for example, new arrivals cannot 'squeeze into' a linear habitat if the spacing between the centres of established territories is less than $4r$, which is to say if the density exceeds $L/4r$ (one-half of that calculated above). If the animals arrive independently and randomly in time, and if each proceeds fairly quickly to phase two, the eventual density will be somewhere between the two extremes just mentioned, and can fluctuate over this factor-of-two range, depending on the statistical accidents of any one pattern of arrival.

These ideas find some support in laboratory studies¹⁰ on territory establishment in sticklebacks. If many male sticklebacks are introduced into a tank simultaneously, the final density of males establishing territories is nearly twice that attained when the same number of males is introduced successively¹⁰. It would be fascinating to carry out similar experiments with male mouthbrooder fish (*Tilapia mossambica*) whose territories in tanks show up beautifully as patterns caused by ridges of

heaped-up sand around the depressions scooped out by the fish (see figure).

This model of territory establishment may not be altogether realistic. In Warner's case, for example, females seem to prefer particular mating sites over others, thus determining the location of male territories. Nevertheless, we should still expect synchrony of territory establishment by males to be important, particularly when a single potential territory

can contain more than one mating site. It would be interesting to see what differences, if any, would ensue if Warner were to repopulate his emptied reefs by sequential rather than simultaneous introductions, or by introducing significantly more or fewer fish than were originally present. □

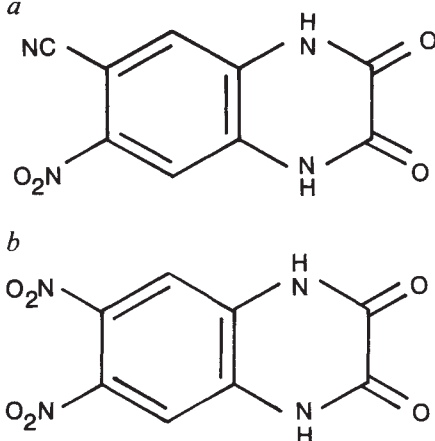
Robert M. May and Paul H. Harvey are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Neurobiology

Quisqualate receptor antagonists

Alan C. Foster

HONORE *et al.* in a recent issue of *Science* describe two new potent antagonists for the quisqualate and kainate subtypes of excitatory amino-acid receptors present on neurons of the central nervous system. Although the three-receptor scheme for excitatory amino-acids is well-established, it is the *N*-methyl-D-aspartate (NMDA) receptor subtype which has received most attention because of the availability of a



The chemical structures of a, CNQX and b, DNQX.

potent and selective antagonists. The discovery of potent antagonists of quisqualate and kainate may now provide the long-awaited tools for detailed investigations of these 'non-NMDA' receptor subtypes. Indeed, results obtained with the new compounds are already challenging the three-receptor scheme and conventional views of excitatory amino-acid function in the central nervous system. But the new compounds 6-cyano-7-nitroquinoline-2, 3-dione (CNQX or FG9065) and 6,7-dinitroquinoline-2,3-dione (DNQX or FG9041) (see figure), may not be the perfect tools because they have additional antagonistic actions at the NMDA receptor which appear to be mediated through an allosteric site.

Honore *et al.* identify the new compounds in radioligand-binding assays using [³H]- α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) and

[³H]kainate as probes for quisqualate and kainate receptors, respectively. With estimated affinities of 0.3–0.5 μ M for the [³H]AMPA binding site and 1.5–2.0 μ M for the [³H]kainate binding site, CNQX and DNQX are two to three orders of magnitude more potent than the previously used non-NMDA receptor antagonists L-glutamate diethylester (GDEE) and γ -glutamylaminomethyl-sulphonate (GAMS)¹. Electrophysiological experiments *in vivo*¹ and *in vitro*^{2–4} confirm that CNQX and DNQX antagonize quisqualate, kainate and AMPA responses in a competitive manner. But quantitative studies *in vitro*^{3–5} reveal that both CNQX and DNQX also block NMDA responses non-competitively^{3,4}. This interaction with the NMDA receptor does not show up in the binding experiments of Honore *et al.*¹ because they used [³H]3-(2-carboxypiperazine-4-yl) propyl-1-phosphonate (CCP), a competitive antagonist which labels the transmitter recognition site of the receptor, as the radioligand. In fact, CNQX and DNQX block NMDA responses by competing with glycine⁶ for an allosteric site within the receptor that modulates receptor activity⁷. This non-competitive antagonism of NMDA receptors is an unwanted 'side-effect' of these compounds, but is mediated via an intriguing interaction with this allosteric site. For practical purposes, the use of CNQX and DNQX as selective non-NMDA receptor antagonists should not be unduly compromised by their NMDA-receptor antagonist properties, provided appropriate control experiments are carried out to check the receptor specificity under the experimental conditions used.

The new work raises some fundamental questions about quisqualate and kainate receptors themselves. CNQX and DNQX are greater than 5 times more potent as inhibitors of [³H]AMPA binding than [³H]kainate binding¹, and yet kainate and quisqualate responses are equally sensitive to antagonism by these compounds^{1,5}. This adds to the growing suspicion that high-affinity [³H]kainate binding sites

1. Krebs, J.R. & Davies, N.B. *An Introduction to Behavioural Ecology* (Blackwell, Oxford, 1987).
2. McNab, B.K. *Am. Nat.* **97**, 133–140 (1963).
3. Schoener, T.W. *Ecology* **49**, 704–726 (1968).
4. Harvey, P.H. & Mace, G.M. *Nature* **305**, 14–15 (1983).
5. Owen-Smith, N. Q. *Rev. Biol.* **52**, 1–38 (1977).
6. Bradbury, J., Gibson, R. & Tsai, I.M. *Anim. Behav.* **34**, 1694–1709 (1986).
7. Krebs, J.R. & Harvey, P.H. *Nature* **333**, 12–13 (1988).
8. Warner, R.R. *Nature* **335**, 719–721 (1988).
9. Maynard Smith, J. *Models in Ecology* (Cambridge University Press, 1974).
10. van den Assem, J. *Behaviour* **16**, 1–164 (1967).
11. Wilson, E.O. *Sociobiology* (Harvard University Press, Cambridge, Massachusetts, 1975).

are in many cases unrelated to depolarizations caused by kainate^{1,8}, and suggests that kainate acts directly on quisqualate receptors to produce these neuronal responses. Thus, CNQX and DNQX could turn out to be true 'quisqualate receptor' antagonists.

It is generally believed that non-NMDA receptors mediate fast neurotransmission at most excitatory synapses in the central nervous system. Experiments using CNQX and DNQX show this to be the case in the hippocampus^{2,9}, but there may be some surprises in store. Honore *et al.*¹ find that synaptic responses in the spinal cord *in vivo* are unaffected by CNQX and DNQX. In addition, it was reported¹⁰ at a recent meeting in Brazil that these compounds are surprisingly poor anticonvulsants, considering that epileptiform activity is thought to be propagated through

neuronal circuits in which non-NMDA receptors play a role. Future studies with CNQX and DNQX, together with analogues possessing a greater degree of selectivity, should help to unravel the identities of non-NMDA receptors and help to elucidate the function of excitatory amino-acid neurotransmitter systems. □

1. Honore, T. *et al. Science* **241**, 701–703 (1988).
2. Blake, J.F. *et al. Neurosci. Lett.* **89**, 182–186 (1988).
3. Birch, P.J. *et al. Eur. J. Pharmac.* **151**, 313–315 (1988).
4. Verdoorn, T.A. & Dingledine, R. *Soc. Neurosci. Abs.* **14**, 480.1 (1988).
5. Drejer, J. & Honore, T. *Neurosci. Lett.* **87**, 104–108 (1988).
6. Birch, P.J. *et al. Eur. J. Pharmac.* (in the press).
7. Johnson, J.W. & Ascher, P. *Nature* **325**, 529–531 (1987).
8. Kemp, J.A. *et al. Br. J. Pharmac.* **91**, 314P (1987).
9. Andreasen, M. *et al. J. Physiol., Lond.* **403**, 57p (1988).
10. Jensen, L.H. & Sheardown, M. *Neurochem. Int.* **12**, Suppl. 1, 8 (1988).

Alan C. Foster is at Merck Sharp & Dohme Research Laboratories, Neuroscience Research Centre, Eastwick Road, Harlow, Essex CM20 2QR, UK.

Supernova 1987A

The pulsar remains hidden

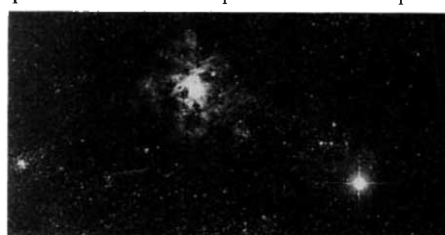
Paul Murdin

THE neutrino flux observed from Supernova 1987A in the Large Magellanic Cloud was in excellent agreement with that predicted from calculations of the formation of a neutron star. Astronomers expect that newly born neutron stars should appear as pulsars, but at a recent meeting*, many expressed worry that no pulsar in SN1987A is yet making its presence known.

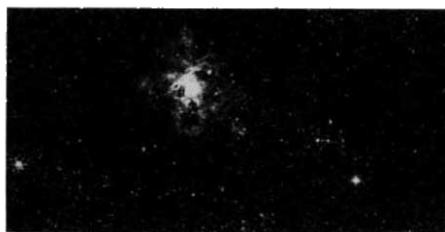
The problem is that all the power feeding the supernova at the present time is from a single source — the radioactive decay of 0.08 solar masses of ⁵⁶Ni (and its β -decay product, ⁵⁶Co), which was formed by nucleosynthesis in the initial explosion. Energy released by β -decay is viewed directly in the form of γ -ray spectral lines from ⁵⁶Co, and indirectly in the form of high-energy X-rays and of optical and infrared radiation, all resulting from the interaction of the γ -rays with the supernova envelope which surrounds the radioactive material. From 120 until 265 days after the explosion, the combined optical and infrared emission from the supernova declined exponentially with a time constant appropriate to the β -decay (see figure). No source of energy with a different time behaviour was apparent.

Up to day 265, the supernova envelope was opaque to X- and γ -rays, but then began to become transparent. Since day 265, the infrared and optical emission has fallen faster than the previous exponential decay and there is a 30 per cent deficit in optical and infrared emission from the supernova against the power input from the decay of 0.08 solar masses of ⁵⁶Co (R.

Catchpole, South African Astronomical Observatory). The missing power in the optical and infrared is, however, well accounted for by the additional power from γ - and X-rays revealed beneath the envelope. To an accuracy of about 5 per cent, radioactivity accounts for all of the supernova output, which sets an upper limit of 5×10^{31} W to the power of any pulsar inside the supernova. This equals



24 April 1987



17 October 1987



4 July 1987



11 February 1988

The declining radiance of SN1987A, observed with the UK Schmidt Telescope.

the power of the pulsar in the Crab Nebula (W. Arnett, Chicago University). If a pulsar is hidden in the supernova, it is certainly not a powerful one and was born rotating slowly, with little rotational kinetic energy to feed into the supernova, or with a weak magnetic field.

If there is a weak pulsar in the supernova, it could be generating soft X-rays by accreting any material which falls

back from the envelope. The maximum power of this process is given by the Eddington limit, beyond which the pressure of the outgoing radiation reverses the accretion flow. This limit is within the observational maximum (R. Chevalier, Virginia University).

Alternatively, the pulsar may be a source of energy for a kind of supernova remnant known as a plerion (F. Pacini, Florence University). The plerion, a nebula formed from the relativistic electrons generated by a pulsar, would add power to the γ -rays produced from radioactive decay. The two sources would be difficult to disentangle because Compton scattering in the supernova envelope degrades the energetic radiation to an output spectrum of hard X-rays, with a sharp cut off below 20 keV and a power-law distribution irrespective of the details of the input spectrum (M. Salvati, Istituto di Astrofisica Spaziale, Frascati).

Attempts to use X- and γ -ray intensities to distinguish pulsar emission (which is constant over periods of a year) from radioactive decay (which falls exponentially) are frustrated by the wide dispersal of radioactive material during nucleosynthesis. This leaves theorists without good constraints on how the material would gradually be revealed as the supernova expands, and models invoking radioactivity alone can be made to fit the intensity curves well.

But, according to Salvati, no matter how the radioactive material was distributed initially, we will certainly be able to see all of it by Christmas, by which time

the supernova will be completely transparent. From this time, the flux of γ -rays and hard X-rays will not depend on details of the structure of the envelope, and the flux must decline exponentially if it comes entirely from radioactivity. Then the presence or absence of a pulsar will be resolved unambiguously. □

Paul Murdin is at the Royal Greenwich Observatory, Hailsham, East Sussex BN27 1RP, UK.

*General Assembly of the International Astronomical Union, Baltimore, 2–11 August 1988.

Exposed continental crust

Deep geology out in the open

John Percival

UNTIL direct observation of the lower 20–30 km of the continental crust becomes feasible through drilling, earth scientists will continue to use surface observations as their primary source of information for the deep crust. Studies of uplifted cross-sections through parts of the crust, using integrated geological, geochemical and geophysical approaches, give critical links between surface geology and geophysical (seismic or electrical) soundings of the Earth at depth. A multidisciplinary group gathered recently* to review and assess the significance of proposed cross-sections from around the globe. In only a few rare examples, such as the study by A.V. Boland *et al.* in this issue (*Nature* 335, 711–713; 1988), have geophysical links been made between the contemporary deep crust and out-of-place surface exposures, most of which are isolated by faults.

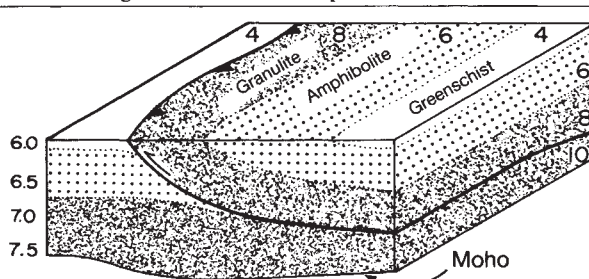
Almost by definition, one expects to see only the surface of continental crust, the deep crust being concealed by the intervening layers of rock. But at several major inclined faults, rooted in the lower crust, rocks have been raised from great depth to the surface, so that one can walk across the raised vertical cross-section (see figure). Because the uplift is rapid, rocks are frozen into the phases appropriate to the temperature and pressure of their original depth, and it is possible to see the history of metamorphism and crustal development at the exposed section.

On approaching the inclined fault, one observes an increase in metamorphic grade, commonly to granulite facies, a decrease in mineral radiometric ages and changes toward more mafic bulk compositions and higher seismic velocities: changes commonly associated with increased depth. Although these criteria have been used successfully to identify crustal sections of Archaean to Cenozoic age, uplift mechanisms vary considerably from the simple model in the figure, to extensional processes in the case of metamorphic core complexes. Most sections have experienced at least two uplift increments.

The meeting included an excursion across the 120-km-wide Kapuskasing uplift and a traverse of the spectacularly exposed Grenville Front near Killarney. Many similarities and differences emerged between the exposed cross-sections (including Ivrea, Calabria, Kapuskasing,

Pikwitonei, Prince Rupert, south India, Tehachapi, Doubtful Sound and Ruby Range). All sections contain magmatic components, and virtually all include supracrustal material (metamorphosed sediments) at the highest metamorphic grade, the deepest level.

A simple explanation of these components involves a common history of oceanic deposition, followed by arc magmatism and incorporation into the



Characteristics of exposed crustal cross-sections. Crustal thickness, 20–50 km; lateral extent of oblique cross-section, 20–150 km. Left-hand scale, seismic velocity (km s^{-1}); top and right-hand scale, palaeopressure (kbar). (After Fountain, D.M. & Salisbury, M.H. *Earth planet. Sci. Lett.* 56, 263–277; 1981.)

margins of stable continental features (cratons), before the eventual uplift. If the sections are representative of the deep crust, then the similarities of lithology in the wide age range sampled suggest that much of the continental crust may record a similar evolution. But an important difference seems to distinguish Archaean from younger terranes. Whereas Cretaceous arcs such as created the Sierra Nevada batholith were active across narrow regions for long periods (J. Saleeby, California Institute of Technology), Archaean magmatism was widespread but short-lived (T. Krogh, Royal Ontario Museum).

Although exposed cross-sections have high-grade rocks at their base, most have not experienced pressures greater than 8–10 kbar, corresponding to a maximum depth of 25–30 km. This is less than the total thickness of the crust, and mid-crustal detachment is indicated, as discussed by Boland *et al.* in this issue.

Structural events with very different upper and lower crustal expressions have been recognized in several sections. Extensional shear zones at depth in the Ivrea zone (K. Brodie, Imperial College, London; M. Handy, University of Bern) are associated with formation of high-level rift basins. Deep-level unroofing of a core complex at Doubtful Sound is correlated with Cretaceous rifting in New Zealand (G. Gibson, Darling Downs Institute). Geochemical depletion of melted pelites (clays) in the Calabria zone may be related

to the formation of higher level, S-type granites (derived from sedimentary material; V. Schenk, University of Kiel).

Crustal structure studies face the problem of an imperfect record. A pseudo-stratigraphic approach to constructing a composite crustal section from depth-calibrated fragments was advocated by D. Fountain (University of Wyoming). Significant steps towards quantifying three-dimensional structure, by linking surface structures to depth, have been made by reflection profiling over high-grade terranes in Canada (A. Green, Geological Survey of Canada). For example, deep crustal reflections, traced to the surface in the Kapuskasing uplift, arise from compositionally layered gneisses.

High-grade terranes help elucidate the nature of the lower crust only if the structural links can be established. Boland *et al.* present results of a study aimed at constraining the subsurface structure of the Kapuskasing uplift by seismic refraction methods. Several models have been proposed to explain this feature, describing it as a suture, a failed rift arm, a transcurrent (strike-slip) fault or an intracratonic thrust (Percival, J.A. & Card, K.D. *Geology* 11, 323–326; 1983). The new refraction work supports the last hypothesis.

Data from a line crossing the uplift demonstrate the presence of material with an unusually high seismic velocity ($6.6\text{--}6.7 \text{ km s}^{-1}$) at shallow depths beneath the high-grade Kapuskasing rocks. The anomaly can be traced, dipping at about 15° to the west, to normal levels for material of this velocity, at about 20 km. The Moho (separating the crust and mantle) deepens from an average of 40 km to as much as 54 km below the uplift. Together, the block of high-velocity material extending towards the surface and the Moho downwarp indicate crustal shortening. The 20-km level may mark a delamination horizon during shortening, separating lighter material that was thrust up and denser material that flowed down to form a root. Crustal delamination (Oxburgh, E.R. *Nature* 239, 202–204; 1972) seems common during shortening in orogenic (mountain-building) belts; the Kapuskasing data could be an intracratonic example.

Identification of the sole of the thrust at 20 km implies that rocks from below this level (fully 20 km of contemporary lower crust) are not represented in the exposed cross-section. Some constraint on the nature of this material is provided by its velocity of $7.2\text{--}7.7 \text{ km s}^{-1}$. Rocks with similar velocities (mafic gneiss, anorthosite), exposed near the base of the exposed section, can be inferred to become more abundant with depth. □

John Percival is in the Geological Survey of Canada, Ottawa K1A 0E8, Canada.

*Exposed cross-sections of the continental crust, NATO Advanced Study Institute, Ontario, 17–27 September 1988.

Quantitative genetics

Character dissection

Richard W. Michelmore and Douglas V. Shaw

Most traits exhibit continuous variation resulting from the action of multiple genes modified by the environment. Determining the location and number of genes that condition such quantitative traits and estimating the magnitude of individual gene effects have occupied quantitative geneticists since early this century^{1,4}. On page 721 of this issue⁵, Paterson, Lander and collaborators demonstrate new analytical methods and use a detailed linkage map to dissect the genetic factors that condition several fruit characters which segregated in an interspecific cross of tomato. In a companion paper⁶, Lander and Botstein provide the theoretical basis for this analysis. These methods confer increased efficiency to the dissection and localization of individual factors conditioning quantitative traits and enhance mendelian approaches to the study of quantitative genetics. This work should ultimately lead to precise biochemical and molecular analyses of the individual genes underlying quantitative characters.

The generation of detailed linkage maps has recently provided adequate numbers of marker loci for extensive dissection and manipulation of quantitative traits. Although these opportunities have been much reviewed, few empirical analyses have been reported. Linkage maps based on RFLP (restriction fragment length polymorphism) loci provide many codominant markers. When large numbers of markers are available, segregating loci can be chosen to mark most regions of the genome. Parents can be selected primarily based on their phenotype for the trait of interest rather than for the presence of distinguishing marker alleles. Lander and Botstein⁶ provide criteria for designing experiments to maximize the chances of detecting quantitative trait loci (QTL), based on the magnitude and number of segregating factors and the relative difference between parental phenotypes compared with the magnitude of environmental variation.

Maximum likelihood

Paterson *et al.* in this issue⁵ employ interval mapping using the method of maximum likelihood. These procedures have been adapted for quantitative traits by Lander and Botstein⁶ from their own methods for analysing complex traits in humans. Interval mapping assesses the effects of each genomic segment, located between pairs of marker loci, rather than the effects associated with individual loci. This reduces the confounding effects of recombination between marker loci and

QTL, efficiently exploits the information from the RFLP linkage data, and provides greater precision than was previously attainable. The authors also chose a stringent threshold for a significant lod score to reduce the probability of false positives. Lod scores refer to the log of the ratio of the odds that there is linkage divided by the odds that there is no linkage.

The results are depicted as 'QTL likelihood maps', a highly informative format to summarize large amounts of information. Likelihood maps represent the probability that one or more QTL lie at a particular point of the genome; this should not be confused with the magnitude of the contribution made by each region. Lander and Botstein also discuss⁶ the increased mapping efficiency that could be achieved by selective genotyping of only the extreme progeny (these individuals tend to provide the most linkage information), by progeny testing to reduce the confounding effects of environmental variance, and by using 'simultaneous search' to reduce genetic noise. These procedures, however, are not reported in this analysis of tomato fruit data by Paterson *et al.* in this issue.

Research in quantitative genetics has thus far relied primarily on biometrical approaches that analyse phenotypic distributions and estimate the statistical effects of genes. This approach has been productive but has been complicated because mathematical models that describe multilocus systems are complex, and broad application of biometrical procedures requires simplifying assumptions of questionable validity. At present, there are at least three areas of research contributing to the development of quantitative genetics: the use of computer simulation for investigation of mathematically complex but realistic models; estimation of, and correction for, bias introduced by violated assumptions of a model; and detailed analyses of individual factors conditioning quantitative traits. The papers by Paterson *et al.* in this issue and by Lander and Botstein make important contributions to the last area.

In common with earlier analyses, there are several limitations with the new methods. Not all quantitative traits will be amenable to dissection. Lander and Botstein conclude that QTL will be most efficiently detected when the parental phenotypes differ by at least k environmental standard deviations, where k equals the number of large effective factors. Analysis of traits conditioned by numerous loci or those with large environmental effects relative to genetic

effects will require unreasonably large experiments. Further, Paterson *et al.* analysed data gathered at one location in one experiment, which provides no information on genotype by environment ($G \times E$) interactions. Plant and animal breeders frequently experience cases where specific genotypes perform differently in different environments. $G \times E$ interactions offer both opportunities and challenges for dissection of QTL. It should be possible to demonstrate the differential actions of different genes in different environments. More importantly, the number of genes that can condition a trait and the size of their effects must be determined using representative sets of environments. This may greatly increase the scale and complexity required for valid experiments. The considerations outlined by Lander and Botstein to increase mapping efficiency, therefore, become all the more important.

Resolution

A third issue concerns resolution. The analysis cannot distinguish between groups of tightly linked loci with small effects and a single locus of large effect. Residual linkage blocks could provide statistically significant associations; this is all the more likely when an interspecific cross is analysed, as in the present case, where incomplete pairing and reduced recombination occurs. Interval mapping can resolve loosely linked QTL as indicated by the analysis of chromosome 10 by Paterson *et al.* Identification and separation of more tightly linked QTL will require sophisticated experiments that use isogenic lines and analyses of recombinants within individual genomic intervals, similar to those used by Shrimpton and Robertson with the fruitfly *Drosophila melanogaster*^{8,9}. Distinguishing between tightly linked loci and single loci is important to both applied and fundamental studies. Selection strategies would differ dramatically depending on whether clusters of genes or single genes are being selected. If genes are to be studied at the biochemical or molecular level, it is critical to determine how many genes are being targeted.

Genetic improvement based on selection of marker genes has been frequently advocated but remains untested for quantitative traits. Biometrical approaches have been exceptionally successful in the absence of precise knowledge about the number or location of the genes involved. Classical selection methods have worked

1. East, E. M. *Genetics* **1**, 164-176 (1915).
2. Wehrhahn, C. & Allard, R. W. *Genetics* **51**, 109-119 (1965).
3. Edwards, M. D., Stuber, C. W. & Wendel, J. F. *Genetics* **116**, 113-125 (1987).
4. Thoday, J. M. *Nature* **191**, 368-370 (1961).
5. Paterson, A. H. *et al.* *Nature* **335**, 721-726 (1988).
6. Lander, E. S. & Botstein, D. *Genetics* (in the press).
7. Lander, E. S. & Botstein, D. *Cold Spring Harbor Symp. quant. Biol.* **51**, 49-62 (1986).
8. Shrimpton, A. E. & Robertson, A. *Genetics* **118**, 437 (1988).
9. Shrimpton, A. E. & Robertson, A. *Genetics* **118**, 445 (1988).

particularly well for improving traits conditioned by few genes of large effect which, as Lander and Botstein discuss, are the traits most amenable to dissection. Valid mapping experiments and marker-based selection are expensive and in many cases may not be justified with the technology currently available. Indirect selection based on marker genes should be evaluated by simultaneously considering selection intensity, heritability, genetic correlation, generation-cycle time, and relative cost of each alternative on a case-by-case basis. Marker-based selection is likely to prove highly effective in specific cases; however careful theoretical consideration of direct and indirect selection alternatives may save considerable time and resources.

Accurate localization of individual QTL generates the potential for reductionist approaches to quantitative

genetics. Isogenic stocks for individual QTL may be developed by selecting for flanking marker alleles. Such isogenic stocks will increase the precision of biochemical and physiological analyses by reducing the genetic and environmental noise. Ultimately, genes may be cloned for traits such as hypertension, atherosclerosis and inherited disorders in experimental animal lines, and for some agriculturally important traits in plants. This, however, is likely to necessitate careful genetic and biochemical characterization of individual QTL using isogenic stocks. As with the analysis of many mendelian traits, careful definition at the classical level is likely to be a prerequisite for successful analysis at the molecular level. □

Richard W. Michelmore and Douglas V. Shaw are in the Departments of Vegetable Crops and Pomology, respectively, University of California, Davis, California 95616, USA.

Solid-state physics

Photons reveal magnetic effects

Stephen Lovesey

X-RAY beams are becoming a favoured probe of the magnetic properties of materials. Traditionally, neutron scattering was the technique preferred by condensed-matter physicists studying magnetic effects in metals and other solids, including the new high-temperature superconductors. X-ray photon factories, using circulating electron beams (synchrotrons) to generate intense, tunable, polarized beams, together with advances in spectrometer design have been the principal spur to the new approach. In contrast, the theories underlying these techniques are well established. Indeed, the theory behind new work by Doon Gibbs and colleagues^{1,2} is contained in the Kramers–Heisenberg dispersion formula, derived in 1925 and pre-dating quantum mechanics. And magnetic photon scattering, used in recent work by Cooper *et al.*³, was predicted in 1954.

X-rays scatter from atoms in materials as a result of interacting with the atoms' constituent electrons, which behave simultaneously as receiving and transmitting antennae, absorbing and re-emitting the photon energy. The interaction occurs, in a quasi-classical picture, by X-rays coupling through their electric field to the electrons' charge, or through their magnetic field to the electrons' charge or magnetic dipole moment (Fig. 1). In the case of electrons bound to atoms in a material, additional constraints apply, the electrons being confined to certain energy states and to orbits that themselves possess a magnetic moment.

In general, charge (or Thomson) scattering dominates all other contributions

and is the interaction that reveals the electron density distribution in crystallography. The magnetic contributions can be enhanced under certain circumstances and it is their effect that is now being exploited in the new studies. In the technique used by Gibbs and colleagues, the X-rays are tuned to a frequency close to that of a transition between allowed electron states in the material. An electron initially in an atomic core level is excited by an incident photon into an intermediate level during the scattering process, through an effect expressed in the Kramers–Heisenberg dispersion formula.

The initial and final states are identical in the Bragg scattering events under consideration. But the sensitivity to magnetic effects is purely quantum mechanical in origin. There is an 'exchange interaction' between a core level state and the intermediate magnetic atomic (or band) state which reflects the Pauli exclusion principle — the states can be notionally swapped. The same type of exchange interaction, operating between atomic magnetic moments, is the origin of the Weiss molecular field that causes magnetic ions to align during a phase transition to an ordered magnetic state. The origin and nature of the magnetic sensitivity is highlighted by the fact that the excitation process in resonant scattering occurs through the electric-field components of the photon, not its magnetic-field component. For some transitions, the resonant exchange scattering is comparable in magnitude to the Thomson scattering. It has a distinct polarization dependence that can be seen using the

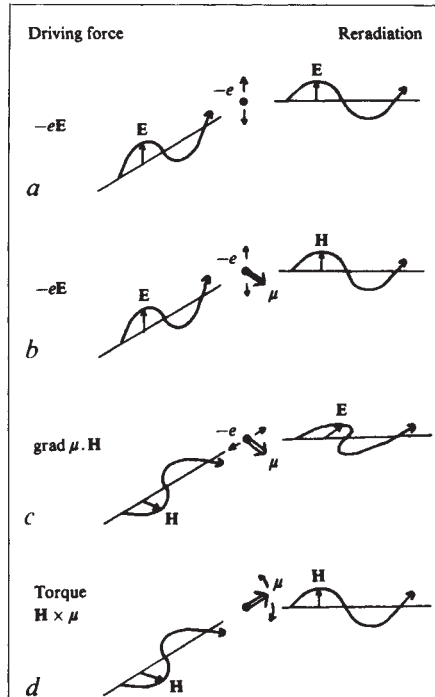


Fig. 1 X-rays can be scattered off electrons by electric or magnetic interactions. *a*, Charge, or Thomson, scattering: the electric field of the X-ray accelerates the electron which re-radiates the X-ray electrically. This is the strongest contribution. *b*, Re-radiation occurs magnetically. *c*, The magnetic field of the X-ray accelerates the electron, which re-radiates electrically. *d*, The field causes the magnetic dipole moment of the electron to precess. (After ref. 8.)

linearly polarized beams readily available at synchrotron sources. In consequence the resonant magnetization-sensitive component is well characterized.

The metal holmium has an unusual anti-ferromagnetic phase between about 20 and 134 K, in which the magnetic moments in successive crystal planes form a spiral staircase. Gibbs and co-workers have previously studied this phase using X-rays¹, confirming earlier neutron-scattering data, that showed that the pitch of the spiral increases as the crystal is cooled (Fig. 2). The spiral repeat length is revealed in magnetic satellite peaks around the principal Bragg peaks.

Gibbs and co-workers now observe² that satellites up to the fourth harmonic are strongly enhanced as the energy of the X-rays is tuned through a $2p$ absorption edge. Each component has a specific polarization signature. These features are consistent² with electric quadrupole and dipole transitions to the $4f$ and $5d$ levels in the metal, respectively, and can be used to probe the magnetic structure of these bands. The strength of scattering depends on the spatial symmetry of the initial and intermediate state, and a departure from spherical symmetry is necessary. Gadolinium has a half-filled, and hence spherical, valence shell, for example, and no magnetic resonant contribution is predicted

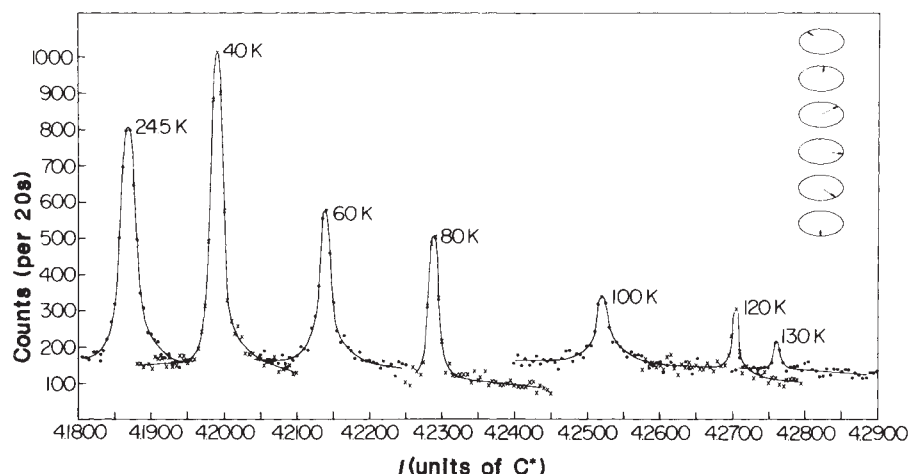


Fig. 2 Magnetic satellite peaks in the diffraction pattern of spiral-phase holmium (inset) are related to the repeat length of the magnetic spiral. As the temperature increases, the repeat length becomes shorter (moves right; the units of abscissa are inverse length). (From ref. 4.)

from the K (innermost-electron) atomic shell, which is also spherical. But there should be an effect for the higher 2*p* core states, which are not. The observed resonances do not seem to occur through relativistic spin-dependent processes advocated previously⁵ and invoked in the interpretation of resonant magnetic effects in ferromagnetic nickel⁶.

There are common features between the direct magnetic interactions of neutrons and photons with unpaired electrons⁷. In particular, both couple to the electron spin and current (orbital) densities and their scattering amplitudes depend on the polarization of the incident beams. The latter effect has been exploited for many years in neutron diffraction, to obtain high-precision maps of the magnetization density throughout crystal unit cells. Changing the magnetic scattering amplitude by changing the polarization of the incident beam provides good signal sensitivity and preserves the state of the target sample. Magnetization density maps supplement independent data, obtained with nuclear-magnetic-resonance and Mössbauer techniques for example, and provide stringent tests of our understanding of highly correlated magnetic states in metals and compounds, including covalency effects.

Whereas the magnetic neutron interaction can be derived from a classical picture of the neutron magnetic moment interacting with the electronic moments and current, the magnetic photon interaction is a relativistic effect and not so readily visualized. A correction to Thomson scattering, it is a relatively weak effect for photon energies small compared with the electron rest mass (511 keV). Even with synchrotron sources, to discriminate it unambiguously in the scattered signal poses problems for the experimentalist. At present, the advantages of magnetic photon scattering over neutron scattering are seen to be the possibility of separating the spin and current

contributions in the amplitude, and its relatively good wave-vector resolution⁸.

Synchrotron radiation is predominantly linearly polarized in the plane of the electron accelerator, but weaker circularly polarized X-rays are emitted at a slight angle to the plane. Cooper *et al.*³ use this diverging emission in their magnetic Compton (or inelastic) scattering experiment on a single crystal of iron. Circular polarization in the incident photon beam induces interference between the charge and magnetic scattering amplitudes, which is analogous to the nuclear-magnetic interference created by polarized neutron beams.

Cooper *et al.*³ measured the polarization of the unpaired electron spins in the sample as a function of the electron momenta, rather than as a function of spatial distribution. There is a clear difference along the three crystallographic directions, [111], [110] and [100], that they investigated. The data complement those obtained by neutron diffraction and agree largely with predictions from standard quantum theories of magnetic materials.

The experiment is possible because of the high intensity of the present generation of synchrotrons which produce usable intensities of out-of-plane X-rays. The next generation of machines, being built at Argonne National Laboratory and Grenoble, should provide a further hundredfold increase in flux, unlikely to be matched by improvements in neutron sources. □

1. Gibbs, D. *et al.* *Phys. Rev. Lett.* **61**, 1241–1244 (1988).
2. Hannon, J.P., Trammell, G.T., Blume, M. & Gibbs, D. *Phys. Rev. Lett.* **61**, 1245–1248 (1988).
3. Cooper, M.J. *et al.* *Nature* **333**, 151–153 (1988).
4. Gibbs, D., Moncton, D.E., D'Amico, K.L., Bohr, J. & Grier, B.H. *Phys. Rev. Lett.* **55**, 234–237 (1985).
5. Blume, M. *J. appl. Phys.* **57**, 3615–5004 (1985).
6. Namikawa, K. *et al.* *J. phys. Soc. Japan* **54**, 4099–5004 (1985).
7. Lovesey, S.W. *J. Phys.* **C20**, 5625–5639 (1987).
8. de Bergevin, F. & Brunel, M. in *Structure and Dynamics of Molecular Systems* Vol. 2 (eds Daudel, R. *et al.*) 69–86 (Reidel, Dordrecht, 1986).

Stephen Lovesey is at the Rutherford Appleton Laboratory, Chilton OX11 0QX, UK.

Daedalus

Attacking the vapours

MOLECULAR sieves are the most subtle of porous materials. Efficiently and selectively, a molecular sieve absorbs those vapour molecules that fit neatly into its internal cavities. Heat it up, and it will let them out again. Imagine, says Daedalus, a disk of molecular sieve with one side hotter than the other. The cold side will absorb specific vapour molecules from the adjacent gas. The hot side, with a much greater vapour-pressure for those molecules, will evolve them again. The cold side will gain molecules, the hot side will lose them, and the disparity will cause a steady internal molecular drift from cold to hot through the porous structure of the disk. The whole thing is an elegant, selective molecular pump with no moving parts, and DREADCO's engineers are exploiting it vigorously.

The simplest applications only require the pumping of trace gases, and need only modest areas of molecular sieve. Thus the currently fashionable 'sick building syndrome' afflicts offices which, to save energy, are inadequately ventilated. Debilitating organic vapours build up from the staff and electrical equipment inside. Radon is another scary atmospheric trace contaminant of many homes and buildings. In both cases, a DREADCO sieve-pump set in an outside wall could neatly, continuously, and selectively absorb and expel the nasties. A healthy and refreshing environment would be ensured without throwing away expensively heated air or causing draughts. Daedalus's ingenious 'enriched greenhouse' is a converse invention. It is partly panelled with molecular sieve elements absorbing carbon dioxide. As its interior warms up in the sunlight, carbon dioxide is automatically sieved from the atmosphere outside and pumped in. Bathed in light, heat and extra carbon dioxide, the pampered plants within photosynthesize with tropical luxuriance.

Bigger gas flows start to demand significant fluxes of heat. For greatest efficiency, a sieve-pump should maintain its temperature difference by pumping heat continuously from the cold side of the sieve to the hot side. The DREADCO team are accordingly developing a 'sandwich structure' of electrically conducting charcoal-based sieve material, faced with porous bismuth telluride thermoelectrodes. A potential difference across the faces pumps heat, and therefore molecules, in the same direction. Volunteers are already blissfully evaluating DREADCO's pilot Pumpsocks[®], which use power from a small battery to expel both heat and water vapour from sweaty feet. A bigger unit, expelling water vapour from a house for air-conditioning and dehumidification, is also feasible. David Jones

Relationships among isolates of HIV

SIR—In their informative analysis of restriction-site diversity among multiple molecular clones of three HIV isolates, M.S. Saag *et al.* (*Nature* **334**, 440; 1988) do not estimate restriction-map distances between isolates in the most appropriate way. The impact on the conclusions is minimal in this case, but we wish to draw attention to this point, not only because with a different data set it could have a considerable effect, but also because a

of each pair of restriction maps (see table).

The weighting has only a small effect on the mean distances in this data set because of the rather low frequency of each restriction map in the isolates. In studies of smaller regions of the HIV genome, however, the commonest type might be much more abundant, leading to a serious bias.

From the table we note that the weighting of the distances leads to a smaller fractional reduction in inter- than intra-group

Inter- and intra-isolate diversities of HIV isolates

	Outgroup*	RJS4	WMF1	WMF3
Outgroup	58.7† 58.7	55.4	54.5	56.5
RJS4	54.5	13.4 12.0	41.9	45.0
WMF1	53.4	40.3	14.1 11.6	24.2
WMF3	56.2	44.1	21.0	12.6 10.7

Above diagonal (bold), means of the pairwise restriction map distances (per cent) as calculated by Saag *et al.*; below diagonal, weighted restriction map distances (per cent) between and within groups.

*Group consisting of isolates HXB2, LAV-MAL, LAV-ELI, ARV2 and WMJ1.

†Mean distance of 55 per cent given in error by Saag *et al.*

comparison between the distances published by Saag *et al.* and the corrected distances points to an interesting aspect of these data.

Although the computation of a pairwise distance matrix is, indeed, a valid way of presenting data of this type, the simple arithmetic mean of the distances between each map (mean difference) is not a correct expression of restriction-site diversity, which is best obtained by weighting each distance by the product of the frequencies

diversity. This arises because the commonest clones in each group are relatively more similar to each other than are the less abundant clones. This may be a first indication of some constraint on the rate of divergence between HIV sequences.

ANDREW LEIGH BROWN

PETER BALFE

Department of Genetics,
University of Edinburgh,
West Mains Road,
Edinburgh EH9 3JN, UK

Calculating just how small a whale can be

SIR—Pearson¹ hypothesized that the lower limit of body size for terrestrial endotherms would be that size at which metabolic heat production, as limited by rates of assimilation, would be balanced by rates of conductive and convective heat loss. We have applied his reasoning to lower limits of body size for aquatic endotherms.

The Cetacea and Sirenia give birth in water^{2,3}, and are the only obligate aquatic endotherms. These mammals did not evolve endothermy in aquatic environments⁴, rather aquatic environments were invaded by terrestrial mammals that were already well-developed endotherms.

Erratum

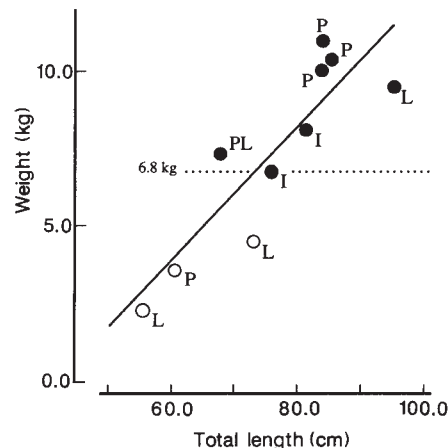
In the letter entitled "Ice-layer dating of eruption at Santorini", by Malcolm K. Hughes (*Nature* **335**, 211–212; 1988), the penultimate paragraph should read "... strongly suggestive circumstantial evidence ..." rather than "... compelling circumstantial evidence ...", an error which was introduced at the editing stage.

Aquatic and terrestrial environments differ in that rates of convective heat loss do not scale directly with differences in heat capacity but are related to differences in Reynold's numbers⁵. In air, at 20°C, the rate of convective heat loss is $3.93 [V/D]^{0.5}$ where V is velocity, and D is a linear measure of size. In water, convective heat loss is much higher $357 [V/D]^{0.5}$. The rate of heat loss in water is 90.8 times greater than in air (the ratio of the coefficients of these equations). Because rates of heat loss scale with body weight (W): $0.224W^{0.57}$ it is possible to calculate the expected minimum size of an aquatic endotherm.

If aquatic endotherms are subject to the same constraints as terrestrial endotherms, the minimum size of an aquatic endotherm would be that size at which the rate of heat loss (RHL) is 90.8 times greater than the heat loss of the smallest (2.5 g) terrestrial endotherm ($RHL_{\text{terrestrial}} = 0.224 [2.5 \text{ g}]^{0.57}$ and $RHL_{\text{aquatic}} = 90.8 (RHL_{\text{terrestrial}})$). Solving $RHL_{\text{a}} = 0.224 [W_{\text{a}}]^{0.57}$ for W_{a} , we

find $W_{\text{a}} = 6.8 \text{ kg}$, the expected minimum weight of an aquatic endotherm.

We recognize that neonates of some terrestrial endotherms weigh less than 2.5 g and their thermoregulatory capabilities are limited. Our prediction, however, is the minimum size at which we might realistically expect endothermic competence. The calculated minimum neonate



Lengths and weights of fetal and postnatal river dolphins⁷. The open circles are fetuses and the filled circles are the smallest postnatal specimens of *Lipotes vexillifer* (L), *Pontoporia blainvillei* (P), *Inia geoffrensis* (I), and *Platanista* spp. (PL). The relationship between length and weight is shown by the least squares regression ($\text{wt (kg)} = 21.1 \text{ length (m)} - 8.66$, $P < 0.001$). The dashed horizontal line is the predicted minimum body mass for an endothermic aquatic neonate.

weight for an aquatic endotherm is 6.8 kg. If an aquatic neonate is smaller than this lower limit, then it cannot generate metabolic heat fast enough to compensate for heat loss, and its body temperature will decline. The predicted value is less than or equal to neonate weights of the smallest aquatic endotherms, the river dolphins (see figure⁷).

JERRY F. DOWNHOWER

Department of Zoology,
Ohio State University,
Columbus,
Ohio 43210, USA

LAWRENCE S. BLUMER

Department of Biology,
Kenyon College,
Gambier,
Ohio 43022, USA

- Pearson, O. P. *Science* **108**, 44 (1948).
- Rice, D. W. in *Orders and Families of Recent Mammals of the World* (eds Anderson, S. & Jones, J. K.) 447–490 (Wiley, New York, 1984).
- Rathbun, G. B. in *Orders and Families of Recent Mammals of the World* (eds Anderson, S. & Jones, J. K.) 537–547 (Wiley, New York, 1984).
- Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
- Gates, D. M. *Biophysical Ecology* (Springer, New York, 1980).
- Bradley, S. R. & Deavers, D. R. *Comp. Biochem. Physiol.* **A 65**, 465–476 (1980).
- Brownell, R. L. in *Reproduction in Whales, Dolphins and Porpoises* (eds Perrin, W. F., Brownell, & DeMaster, D. P.) 149–158 (Reports of the International Whaling Commission, Special Issue 6, Cambridge, 1984).

Pollution by toxic metals on agricultural soils

SIR—We are concerned with the recent suggestion of Peter Moore¹ that there could be recuperative strategies for continuing agriculture on polluted soils rather than focusing on preventing the contamination of sewage sludge at source. Although use of the less contaminated sewage sludges produced in most rural areas will lead to few problems with metals, the difficulty is that most of the sludge to be disposed of each year comes from conurbations and is invariably contaminated with zinc, cadmium, copper, nickel, chromium and lead. Only with more stringent control over effluents from industry could the problem of sludges from cities be substantially reduced.

Metals are not removed from soil by normal cropping nor are they leached from near-neutral topsoils. A maximum of only 0.5 per cent of the amount of each of six metals accumulated in a sludge-treated soil is removed after 20 years of harvesting². For all practical purposes, once metals are added to soil they remain there for thousands of years.

Certainly in the case of elements such as copper, little is absorbed from the soil by most plants unless very high concentrations are present, and as Moore suggests sheathing mycorrhizas are important in preventing metal uptake in very heavily contaminated soils. But not all potentially toxic elements behave in the same way; in particular cadmium is more readily absorbed from the soil and therefore is the element of greatest concern in food chains. There is no evidence of reduced cadmium uptake by plants infected with mycorrhizas. Furthermore, vesicular-arbuscular mycorrhizas, which infect major crop plants, actually increase metal uptake by plants³. Some other crop species are completely non-mycorrhizal, so the proposed¹ protection mechanism could not operate.

It is unlikely that the microbial population of the soil will be the sole solution to avoidance of metal toxicities. We have shown⁴ that toxic effects of metals on *Rhizobium* caused complete suppression of symbiotic nitrogen fixation and reduced yields of clover by 40 per cent in the field. It appears that high metal concentrations cause loss of symbiotic plasmids in *Rhizobium*, leading to a lack of ability to nodulate the host plant⁵. Heavy-metal contamination may also halve the size of the

microbial biomass⁶ and prevents growth and nitrogen fixation by blue-green algae⁷. These effects occur at concentrations of metals close to the European limits⁸.

Because toxic metals will persist for so long, the problem is not transient and could have widespread implications. This is particularly true for land taken out of agriculture. To maintain plant cover, such land will be dependent on biological fixation of nitrogen and cycling of nutrients through the microbial biomass. Surely there is a case for increased stringency in the guidelines for environmental protection, rather than advocating the

increased application to land of sewage sludges contaminated by industry, until we know more about long-term effects of metals on microbial processes in the soil.

KEN E. GILLER

Department of Biochemistry and Biological Sciences,
Wye College, University of London,
Wye, Ashford, Kent TN25 5AH, UK

STEVE P. MCGRATH

Soils and Crop Production Division,
AFRC Institute for Arable Crops
Research,
Rothamsted Experimental Station,
Harpenden, Hertfordshire AL5 2JQ, UK

Estimating the Maunder minimum from auroral data

SIR—Some of the recent papers concerning the solar activity in the seventeenth and eighteenth centuries infer that auroral observations can be used to provide new evidence for the solar minimum described by Maunder¹. For example, Eddy^{2,3} has suggested that auroral data may be used to provide a conclusive proof of this particular solar period.

An examination of historical sources, however, shows that such a premise cannot be accepted. Although the catalogue of Fritz⁴ is certainly invaluable, many other sources must be considered⁵.

A collection of available auroral observations during the second half of the seventeenth century is given in the table. It shows that there is insufficient evidence for a minimum of auroral activity during 1645–1715. There are only a few years during which no aurorae were recorded, and some of these fall into times of solar minimum, when scarcely any aurorae at all are seen in middle latitudes. Therefore

the absence of auroral observations during 1645–1715 is not conclusive evidence for a general minimum of solar activity during this period. Furthermore, the frequency of aurorae during 1645–1715 shows the same characteristics as those seen in other long-term observations from middle latitudes (50–55°N). From records for 1882–1965 the average number of aurorae every year is one to five, depending on the solar activity and solar cycle⁶, and also on the attention paid to them by observers. This is also consistent with the period of the Maunder minimum (see table).

W. SCHRÖDER

Hechelstrasse 8,
D-2820 Bremen-Roenebeck, FRG

1. Maunder, W.E. *Knowledge* 17, 173 (1894).
2. Eddy, J.A. *Science* 192, 1189 (1976).
3. Eddy, J.A. *Sky Telesc.* 517 (1982).
4. Fritz, H. *Verzeichniss beobachteter Polarlichter* (Gerolds, Vienna, 1873).
5. Schröder, W. *Das Phänomen des Polarlichts* (Wissenschaftliche Buchgesellschaft, Darmstadt 1984).
6. Eddy, J.A. *Sol. Phys.* 89, 195 (1983).

Auroral data in central Europe 1645–1707

Year	Number of aurorae	Year	Number of aurorae	Year	Number of aurorae
1645	2	1666	3	1687	1 (2?)
1646	3 (4?)	1667	0	1688	1?
1647	0	1668	1?	1689	0
1648	5 (6)	1669	0	1690	?
1649	2	1670	2 (3?)	1691	0
1650	1	1671	1	1692	1 (3?)
1651	2	1672	1 (2?)	1693	?
1652	1 (3?)	1673	1	1694	?
1653	2	1674	0	1695	2 (3?)
1654	4	1675	0	1696	1 (3?)
1655	1	1676	3	1697	1
1656	0	1677	3	1698	?
1657	4	1678	1	1699	?
1658	2	1679	0	1700	0
1659	0	1680	3 (4?)	1701	0
1660	2	1681	3	1702	2
1661	4	1682	4 (5?)	1703	0
1662	3	1683	2 (3?)	1704	4
1663	3	1684	1	1705	2
1664	3	1685	1	1706	1?
1665	4	1686	5	1707	15

1. Moore, P.D. *Nature* 333, 706 (1988).
2. McGrath, S.P. in *Pollutant Transport and Fate in Ecosystems* (eds Coughtrey, P.J. et al.) 301–317 (Blackwell, Oxford, 1987).
3. Kilham, K. & Firestone, M.K. *Pl. Soil* 72, 39 (1983).
4. McGrath, S.P. et al. *Soil Biol. Biochem.* (in the press).
5. Casella, S. et al. *FEMS Microbiol. Lett.* 49, 343 (1988).
6. Brookes, P.C. & McGrath, S.P. *J. Soil Sci.* 35, 341 (1984).
7. Brookes, P.C. et al. *Soil Biol. Biochem.* 18, 345 (1986).
8. Council for European Communities *J. Eur. Commun.* L181 Annex 1A (1986).

Hot from the laboratory

Robert M. Hazen

Publishers of newsletters have reacted swiftly to meet the demand for information on developments in high-temperature superconductivity. What have these publications got to offer?

THE 1986 revelation of high-temperature superconductivity was unique in the history of materials science, and of publishing. Journals quickly altered their review procedures, and even changed formats, in order to accommodate the ensuing flood of reports and the need for speed. New journals devoted exclusively to high-temperature superconductivity appeared along with dozens of special high- T_c volumes and conference reports. But the normal scientific media have proved too slow for superconductor research circles. In this frenetically competitive atmosphere, knowledge is the most prized commodity, and researchers will gladly pay to get it.

Enter the superconductivity newsletters. Has a publishing niche ever been filled so quickly? Within six weeks of the 'Woodstock of Physics' on 18 March 1987 (the last-minute, all-night superconductivity session at the American Physical Society meeting in New York), the first newsletter rolled off the presses. By the end of the summer eight more had been launched. The race for publishing dollars has paralleled that for higher critical temperatures.

This article reviews most of the newsletter-style periodicals devoted to superconductor science and technology. These publications are distinguished by inclusion of news stories, review articles, lists of work not yet published or behind-the-scenes reports of superconductor research laboratories (for which read gossip), rather than original scientific reports.

The nine titles surveyed are summarized in the table overleaf, which gives the

Supercurrents



The Superconductivity Magazine

vital statistics and brief comments on each of them. Taken as a group, these publications serve a useful — perhaps vital — role by providing a rapid international communications medium. But no one needs to scan more than two or three of them. The question is, which ones?

Without doubt first prize goes to *High- T_c Update*, published by the United States Department of Energy under contract with the Ames Laboratory, Iowa State University. This 8–16-page biweekly newsletter has become the international clearing house for superconductivity preprints. In addition to listing as many as 100 manuscript titles per issue, it features

news stories on the most significant advances and provides space for meeting and product announcements. Best of all, *High- T_c Update* is free! All superconductor researchers should be on its mailing list — the newsletter is available electronically as well as in printed form — and should send their preprints (and thanks) to the editor, Ellen O. Feinberg.

Of those you have to pay for, I found the Washington-based *Superconductor Week* to be the best bet for researchers. This eight-page, weekly publication is devoted exclusively to developments in superconductivity and stands out among commercial newsletters by providing not



only rapid but usually cautious coverage of both high- T_c science and technology. This was one of the first publications to report on the record-setting bismuth and thallium superconductors early in 1988, for example, and its news stories have appeared as formal citations in major physics journals. *Superconductor Week* carefully labels unconfirmed rumours as such, and the editorial staff do their homework, attending the main scientific meetings and maintaining close contact with laboratories around the world. The articles are written in a readable yet no-nonsense journalistic style, though I find the habit of treating strings of individual sentences as separate paragraphs rather distracting. I understand that it comes in electronic form, on NewsNet.

Superconductivity News (not to be confused with *Superconductor News*) is a close second, with particularly strong reporting of high- T_c conferences. A thorough job is usually done in digging out news and analysing the scientific and technological implications. Monthly issues are supplemented by news updates for important stories. The regular "Handbook" feature provides a comprehensive listing of superconductivity events and products in a concise and efficient format. Subediting is slack at times, and the writing style is a bit too cute for my taste, with dozens of self references to "SN" in every issue. Remarks such as "SN hopes researchers will invent an elevated

transition temperature superconductor made from commonly available elements" are space fillers, not news.

New Technology Week, with good coverage of commercial aspects of superconductors and other electronic materials, is another option. Here the emphasis is more on government and corporate politics, rather than the science of materials. In most of the weekly 12-page issues two or more pages of superconductor news are integrated with other features. Competition between the United States and Japan, industrial research strategies and relevant Congressional legislation are typical fare.

Superconductor Week

The overall tone is rather pessimistic, with headlines such as "National Labs Struggle with Technology Transfer" and "U.S. Ceramics Effort Pulling up the Rear". But if materials policy is your concern then *New Technology Week* is a valuable resource. Both *Superconductor Week* and *New Technology Week* go to press late on Friday and appear on Monday morning, thus ensuring the fastest possible coverage of developments.

Of the remaining periodicals, my favourite is *Supercurrents*. Launched in January 1988, this glossy monthly is the most stylish of the lot. Donn Forbes, the editor, seems determined to produce a publication that combines timely features with lasting archival value. Of special interest are interviews with leading researchers (recent issues have featured Art Sleight, Shoji Tanaka and Brian Maple, for example). Well-illustrated review articles on the science of superconductivity and realistic assessments of potential applications add to the non-ephemeral aura. *Supercurrents* will never be the first to report on high- T_c breakthroughs, but it is a publication that materials research libraries should seriously consider adding to their acquisitions list.

It will be years before superconductivity researchers can settle down to a more sedate and normal routine. Meanwhile, armed with these publications, they have a reasonable hope of staying informed. □

Robert M. Hazen is in the Geophysical Laboratory, Carnegie Institution of Washington, 2801 Upton Street, NW, Washington, DC 20008, USA. He is author of *The Breakthrough: The Race for the Superconductor (Summit, 1988)*, in Britain published by Unwin Hyman under the title *Superconductors: The Breakthrough*.

• Just launched by World Scientific is the first issue of *HT_c News*, a free monthly newsletter which will "tend toward a more 'academic' approach", and which contains abstracts of published papers, titles and authors of forthcoming papers, conference notices and "resources information". In the United States World Scientific is at 687 Hartwell Street, Teaneck, New Jersey 07666; in Britain at PO Box 379, London N12 7JS.

Hot from the laboratory — publication details for 1988

Publication and launch date	Publisher	Issues per annum	Subscription per annum	Comments
The Cambridge Report on Superconductivity (Sept. 1987)	Rich-Gumpert Inc. 1 Kendall Square Cambridge Massachusetts 02139	12	\$245 (US), \$270 (elsewhere)	Neither timely nor cost-effective at eight pages per issue. Good coverage of the Cambridge, Massachusetts, scene, but weak on superconductor science and foreign developments.
High-T_c Update (May 1987)	Ames Laboratory/ US Dept of Energy 12 Physics Iowa State University Ames, Iowa 50011	24	Free	The Department of Energy and the editor deserve medals for this vital information clearing house. Proves you can get something for nothing — though foreign subscribers have to agree to provide preprints etc. in return.
Lets Levitate (Aug. 1987)	The Superconductor Hobby Club 1915 Zachary Drive Salt Lake City Utah 84116	6	\$30 (US)	This modest (some might say crude) six-page photocopied newsletter is written for amateur scientists. Short features on the history, applications and recent developments in high- T_c materials are competent and low key, but they don't provide enough of an overview to justify the subscription cost.
New Technology Week (June 1987)	King Communications Group 627 National Press Building Washington, DC 20045	50	\$395	Timely coverage of corporate and government policies and politics affecting materials research and development. Superconductivity is prominently featured in the context of electronic materials. Not the best high- T_c coverage, but valuable for its overview of a variety of new materials.
Superconductivity Flash Report (Sept. 1987)	Flash Report Publishers Inc. Suite 626 Doral Plaza 155 N. Michigan Ave Chicago, Illinois 60601	24	\$345 (US), \$390 (elsewhere)	Each eight-page issue contains broad international news coverage of superconductor science and applications, but most reports are of the snippet variety. Useful as a starting point for tracking down advances in industry.
Superconductivity News (July 1987)	Superconductivity Publications Inc. Suite 2000, 65 Jackson Drive Cranford New Jersey 07016	12	\$300 (US), \$360 (elsewhere)	Twenty-page issues contain a satisfying blend of science and applications, written in an occasionally distracting, light and colloquial style. The best of the non-weekly newsletters, but not as up-to-the-minute as <i>Superconductor Week</i> .
Superconductor News (Aug. 1987)	Superconductor Applications Association 24781 Camino Villa Avenue El Toro California 92630	6	\$45 (US)	Publication of a trade association. Recent eight-page issues contain paragraph-long summaries of work at various industry and university laboratories, but little that is new or of substance. Two-page advertisements for conferences and services administered by the association are a prominent feature.
Superconductor Week (July 1987)	Atlantic Information Services 1050 17th St N.W. Suite 480 Washington, DC 20036	48	\$337 (US, pre-paid); \$357 (US, not pre-paid); add \$74 for air mail elsewhere	Best of breed. Coverage of high- T_c science and technology is rapid and conservative. Editor David Chaffee maintains an effective news balance between laboratory discoveries, corporate developments, and government policy on superconductivity.
Supercurrents (Jan. 1988)	Supercurrents PO Box 889 Belmont California 94002	12	\$60 (US)	This stylish monthly reports important high- T_c developments, along with features of lasting archival value. Interviews with leading researchers, combined with high-quality photographs, sets <i>Supercurrents</i> apart from the newsletters.

Omitted from this round-up are several closely related newsletters on materials science more generally, including *Materials and Processing Report*, *Japan Materials News*, *High-Tech Materials Alert* and *Nikkei High Tech Report*. Also excluded are three failed newsletters — *Superconductor Advisory Newsletter*, *Superconductivity Research, Development and Commercialization Report* and the ambitious and stylish monthly, *SuperConductor World Report*.

Breast feeding, birth spacing and their effects on child survival

Shyam Thapa, Roger V. Short & Malcolm Potts

It has long been known that breast feeding inhibits female fertility and that it is a factor in restricting population growth. But just how important is it?

DURING the course of human evolution breast feeding has come to play a key role in regulating both the mother's fertility and the infant's wellbeing, and we tamper with it at our peril. The contraceptive effect of breast feeding has been recognized since antiquity. Aristotle¹ pointed out that "while women are suckling children menstruation does not occur according to nature, nor do they conceive; if they do conceive, the milk dries up", and modern physiologists have confirmed the truth of this statement. Demographers have shown that the interval between successive births is a principal determinant of marital fertility, and in populations without access to modern forms of contraception this birth interval will be largely determined by the duration of breast feeding^{2,3}. The relationship between changing patterns of breast feeding and the need to adopt artificial contraception in order to achieve optimal birth intervals raises pressing questions for policy makers and moralists alike.

Lactational infertility

Lactational infertility takes two forms. There is a period of complete infertility during lactational amenorrhoea when ovulation is suppressed, and this is followed by a variable period of lowered fecundity after the resumption of ovulatory menstrual cycles. Breast-feeding women who resume menstruation within six months of childbirth will usually do so before the first postpartum ovulation, the first menstruation in such cases being the result of an oestrogen withdrawal bleed as waves of follicular development and atresia occur in the ovaries. Those who remain amenorrhoeic for more than six months are more likely to ovulate before they first menstruate⁴⁻⁶.

Endocrinologists are uncovering the mechanisms by which lactation inhibits reproduction. Afferent neural inputs to the hypothalamus following nipple stimulation seem to cause a local release of beta-endorphin, which in turn inhibits hypothalamic secretion of gonadotrophin-releasing hormone and dopamine, thereby suppressing gonadotrophin secretion and ovarian activity while stimulating the secretion of prolactin^{4,7}.

The key to the short-term and long-term success of lactation as a contraceptive is therefore the frequency with which afferent neural inputs generated by the baby's stimulation of the mother's nipple reach the hypothalamus⁴⁻⁶. Not only do these neural inputs regulate fertility, but elevated prolactin levels stimulate the long-term synthesis and secretion of milk in the breast, while the sucking-induced discharge of oxytocin from the posterior pituitary causes contraction of the myo-epithelial cells around the mammary alveoli, leading to milk ejection. If the alveoli are not repeatedly emptied in this manner, milk production will be inhibited, and eventually lactogenesis will cease altogether⁸. At weaning, or following death of the child, the alveoli are not emptied, milk synthesis ceases and fertility returns. Similarly, any practice such as the use of 'dummies' or 'comforters' that reduces sucking frequency will ultimately

decrease milk production and erode the natural contraceptive effect of breast feeding.

Sucking frequency

Conversely, we know in animals that if the supply of milk initially fails to meet the infant's needs, for example as a result of maternal malnutrition, the infant will demand to be fed more often and the increase in sucking frequency not only helps increase milk production but may also provide additional contraceptive protection⁹. Thus, in normal circumstances, lactation has been nature's contraceptive, but we have unwittingly interfered with this effect as a result of changing breast-feeding practices.

Breast milk is also a potent guardian of infant health. The maternal antibodies found in human milk protect the baby from gastroenteritis and respiratory infections¹⁰. The enteromammary circulation



Mary Evans

History clearly tells the story of the adverse effect of abandoning breast feeding on birth spacing and child survival. A seventeenth-century engraving of Sir Thomas and Lady Remington of Lund, Yorkshire, with their 20 children, five of whom died in early childhood. The skull at the bottom left presumably represents an additional still birth. Lady Remington cannot have breastfed her children; had she done so she could never have produced so many. She probably sent them all out to be wet nursed, as was the custom for the nobility at that time⁴⁵.

ensures that the appropriate immunoglobulin A is secreted into the breast milk within hours of exposure of the mother's gut-associated lymphoid tissue to a potential pathogen^{11,12}. Milk substitutes lack this natural immunological protection, and often exacerbate the situation by being bacteriologically contaminated and incorrectly formulated. Although millions of human babies have been reared successfully without so much as a taste of breast milk, it is still true to say that the artificial feeding of our infants has been the largest uncontrolled clinical experiment in human history¹³. When incorrectly formulated or misused, it has killed, and continues to kill, large numbers of babies in developed and developing countries.

Feeding duration

Using data from various populations, Bongaarts^{3,14} demonstrated that the duration of breast feeding explains 96% of the variance in the duration of postpartum amenorrhoea. His model indicates that the duration of postpartum amenorrhoea is not much increased by lactation beyond 24–25 months, nor substantially decreased when the average time of breast feeding falls below 4–5 months; the greatest effect is found when the duration of lactation varies between 6 and 18 months.

Several studies^{15–17} have documented variations in the duration of breast feeding in different countries. Breast-feeding duration is generally shorter among young, affluent, urban, educated women than in their older, poorer, rural, less-educated counterparts. Declines especially in duration, but to a lesser extent in the incidence of breast feeding, have been reported in several countries in Asia and Latin America^{18–20}, although in a few countries, notably Singapore, Malaysia and Thailand, the situation appears to be stabilizing¹⁹.

Unfortunately, the trends of declining breast feeding and rising prevalence of contraceptive use are not always in step; sometimes, lactational infertility declines more rapidly, as in some subnational populations^{21,22}. Hence the interplay of lactational infertility and contraceptive use must be considered before their ultimate effect on fertility can be assessed.

We have analysed data from the World

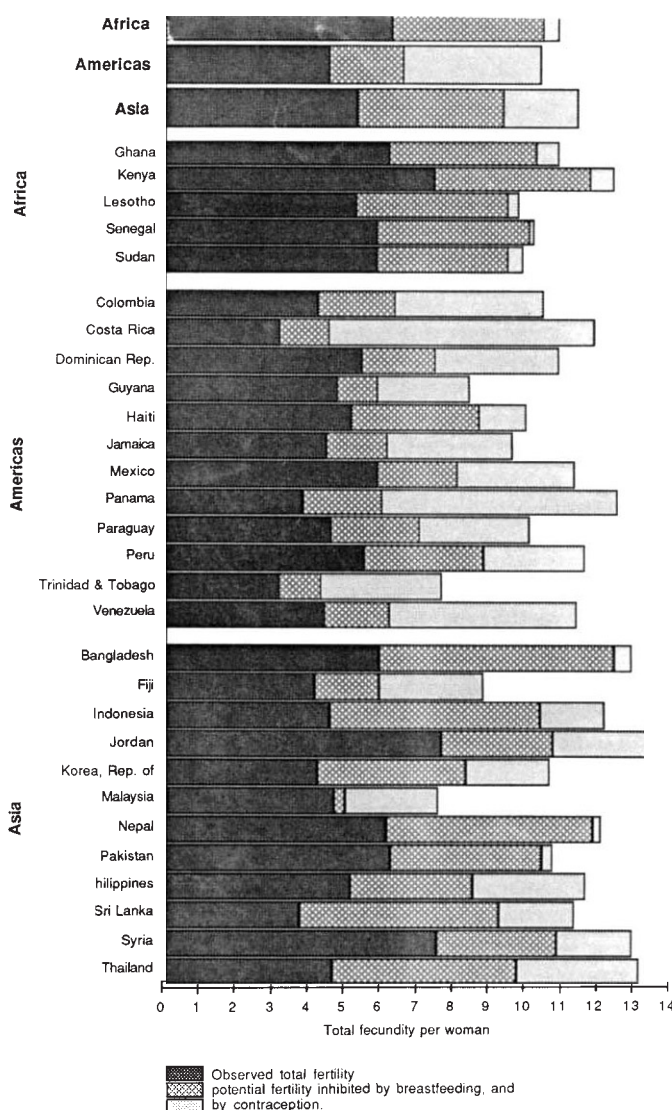


Fig. 1 Total fertility rates in 29 developing countries (defined as the average number of children who would be born to a woman during her lifetime if she were to live through all her childbearing years and conform to the current age-specific fertility rates), and the number of potential births currently inhibited by breast feeding, and by modern forms of contraception. The total length of each bar corresponds to the total fecundity rate per woman.

Fertility Survey (WFS)^{23,24}, the largest social survey ever undertaken, and used it to examine the effect of changing breast-feeding practices on fertility. The data came from approximately 150,000 women of reproductive age (15–49) in 29 developing countries in which live nearly 30% of the total population of the developing world, excluding China, and included nations with high to moderate fertility. The data were collected between 1974 and 1984.

Figure 1 shows the observed total fertility rate per woman, and the potential fertility currently inhibited by breast feeding, and by contraception, in each of the 29 countries. This was calculated by applying the model of the proximate determinants of fertility developed by Bongaarts³ to the WFS data. The model

of proximate determinants is parsimonious, yet it is still the most robust quantitative method with which to assess the relative roles played by various factors influencing fertility. Bongaarts demonstrated that of the seven proximate proportions married, postpartum infecundity, contraception, induced abortion, spontaneous intrauterine mortality, waiting time to conception and permanent sterility—only the first four vary significantly across population subgroups. While early marriage and postpartum infecundity are the principal determinants of fertility in traditional societies, contraception and induced abortion are the principal determinants in more modern, developed societies³.

Figure 1 shows that in Africa the amount of potential fertility reduced by contraceptive use is low, whereas breast feeding inhibits an average of four births per woman. In Asia, breast feeding remains important for fertility control (for example, in Bangladesh, with a population of about 100 million, breast feeding inhibits an average of 6.5 children per woman) but, overall, contraception is four times more important than in Africa. In Central and South America, contraception plays the predominant role and only about one-fifth of the total fecundity reduction is due to breast feeding, although in Peru and Haiti breast feeding still plays a dominant role.

To assess the magnitude of the fertility-inhibiting role of breast feeding, we calculated the expected changes in fertility for the same countries, assuming that the duration of breast feeding were to decline by 25% or 50% and contraceptive usage was to remain unchanged (Table 1). If there are no changes in marriage patterns or the incidence of induced abortion, the only way to offset the potential increase in fertility as a consequence of shortening the duration of breast feeding is to increase contraceptive use.

Contraceptive prevalence

The increased contraceptive prevalence required to offset the projected declines in breast feeding (Table 2) is large in Africa (for example, a 25% breast-feeding decline in Senegal would require almost a tripling in contraceptive prevalence),

more modest in Asia (a 38% increase in current contraceptive use in Indonesia for a 25% fall in breast-feeding duration), and generally small in Central and South America (a 7% increase in contraceptive prevalence in Mexico would be needed to compensate for a 25% fall in breast-feeding duration). Of course, the converse is also true; a rise in the duration of breast feeding among Latin American women would lead to further reductions in fertility, even without any change in contraceptive use.

Child survival

It is a commonsense argument that child survival is a powerful motivational factor leading to a desire for smaller families—but one that assumes enormous importance in both policy formulation and programme development. The WFS data show that a child's risk of dying in the first years of life depends on the preceding birth interval; mortality can be greatly reduced if it is born at least two years after its elder sibling, especially in places where most mothers are uneducated and lack access to primary health-care facilities and sanitation²⁵⁻³⁰. The effect of preceding birth interval on child survival is found to be even greater than the effect of declines in parity or the mother's age at childbirth. Furthermore, the life-saving effects of increased spacing of the preceding birth remains strong, even when the influences of length of previous gestation, survival status of the previous child and maternal educational attainment are taken into account. It has been estimated that a preceding birth interval of less than two years can raise the average chances of a child dying before age five by about 50% (ref. 28). The underlying mechanisms responsible for this effect are obscure, although an ill-defined "maternal depletion syndrome" is often invoked^{29,31}, which could result in the birth of an underweight baby with diminished chances of survival.

There is also a devastating effect of too short a subsequent birth interval on child survival²⁸. As Aristotle pointed out, a new pregnancy cuts off the milk supply, leading to the abrupt weaning of the older child, and this can have disastrous consequences. *Kwashiorkor*, a West African term meaning literally "the disease of the displaced one", is used to describe the symptoms of gross protein malnutrition in the older child if it is suddenly deprived of its principal source of dietary protein through early weaning. In calculating the effect of subsequent birth interval on child survival it is important to avoid overestimates through reverse causation; premature death of the older child for whatever reason may result in a shortened birth interval by hastening the mother's return to fertility. The best estimate of the true influence is that a second birth within 12 months of the index birth raises the risk

Table 1 Current total fertility rate (TFR)* per woman and changes in TFR associated with the decline in the duration of breast feeding by 25% (projection I) and by 50% (projection II)

Region/country	Current TFR	Increase in TFR		Percentage increase in TFR	
		Projection I	Projection II	Projection I	Projection II
Africa	Ghana	6.2	7.0	12.9	27.4
	Kenya	7.4	8.3	12.2	24.3
	Lesotho	5.3	5.9	11.3	26.4
	Senegal	6.9	7.7	11.6	27.5
	Sudan	5.9	6.7	13.6	25.4
	Regional average†	6.3	7.1	12.7	27.0
South & Central America	Colombia	4.3	4.4	2.3	9.3
	Costa Rica	3.2	3.2	0.0	3.1
	Dominican Rep.	5.4	5.6	3.7	9.3
	Guyana	4.7	4.8	2.1	6.4
	Haiti	5.2	5.6	7.7	17.3
	Jamaica	4.5	4.8	6.7	8.9
	Mexico	5.9	6.2	5.1	9.2
	Panama	3.8	4.1	7.9	10.5
	Paraguay	4.6	4.9	6.5	9.8
	Peru	5.4	5.7	5.6	14.8
Asia & Pacific	Trinidad & Tobago	3.2	3.2	0.0	3.1
	Venezuela	4.4	4.6	4.5	6.8
	Regional average†	4.6	4.8	4.3	8.7
	Bangladesh	6.0	6.8	13.3	33.3
	Fiji	4.1	4.3	4.9	9.8
	Indonesia	4.5	5.2	15.6	37.8
	Jordan	7.6	8.1	6.6	13.2
	Korea, Rep. of	4.2	4.7	11.9	28.6
	Malaysia	4.6	4.7	2.2	6.5
	Nepal	6.1	7.1	16.4	36.1
	Pakistan	6.2	7.0	12.9	27.4
	Philippines	5.1	5.5	7.8	15.7
	Sri Lanka	3.7	4.2	13.5	32.4
	Syria	7.5	8.3	10.6	13.3
	Thailand	4.6	5.1	10.9	23.9
	Regional average†	5.4	5.9	9.3	22.2

* Data on current TFR from ref. 43. † Regional averages are unweighted.

Table 2 Contraceptive prevalence required to offset fertility increase due to decline in the duration of breast feeding by 25% (projection I) and 50% (projection II)

Region/country	Current contraceptive prevalence*	Contraceptive prevalence required		Percentage increase in contraceptive prevalence required	
		Projection I	Projection II	Projection I	Projection II
Africa	Ghana	10	19	90	190
	Kenya	9	19	111	200
	Lesotho	5	14	180	400
	Senegal	5	14	180	400
	Sudan	8	19	138	238
	Regional average†	7	17	143	286
South & Central America	Colombia	43	45	5	7
	Costa Rica	64	64	0	2
	Dominican Rep.	32	34	6	19
	Guyana	32	33	3	13
	Haiti	25	27	8	40
	Jamaica	39	42	8	10
	Mexico	30	32	7	20
	Panama	54	58	7	9
	Paraguay	47	51	9	15
	Peru	31	31	0	16
Asia & Pacific	Trinidad & Tobago	60	60	0	3
	Venezuela	60	62	3	5
	Regional average†	43	45	5	12
	Bangladesh	8	19	138	325
	Fiji	41	43	5	12
	Indonesia	26	36	38	77
	Jordan	25	28	12	32
	Korea, Rep. of	35	40	14	40
	Malaysia	33	35	6	15
	Nepal	2	16	700	1,350
	Pakistan	5	16	220	420
	Philippines	36	37	3	14
	Sri Lanka	32	39	22	50
	Syria	30	35	17	23
	Thailand	33	41	24	42
	Regional average†	26	32	23	54

* Contraceptive prevalence refers to the mean of age-specific proportions of women currently using a method of contraception. Data on current contraceptive prevalence from ref. 44. † Regional averages are unweighted.

of dying between the ages of one and five by at least 77%; if the mother is pregnant again by the index child's second birthday, this raises its risk of dying before the age of five by 55% (ref. 28).

It is tempting to make predictions about how many children's lives could be saved if mothers could space their pregnancies by an average of at least two years. For the 29 countries considered here, the current annual total of 2.6 million deaths of children under one year of age might be expected to fall by at least 20% resulting in a saving of about half a million lives a year. Even though Bongaarts³² has questioned whether the provision of family-planning services does in fact reduce infant mortality, Trussell³³ has pointed out weaknesses in his argument, which obscures the very real beneficial effect of family planning in reducing infant and child mortality.

Despite the overwhelming evidence of increased child survival with increased birth spacing, many children in developing countries are still born too closely spaced. This figure ranges from nearly 30% in Asia and the Pacific to 40% in the Americas^{25,34}. Even in sub-Saharan Africa, where traditional child-spacing practices, such as taboos on intercourse during lactation, are more prevalent than in the other two continents, about 20% of babies are still born within two years of a sibling.

Postpartum contraception

There is clearly an urgent need to design new forms of postpartum contraception for developing countries that do not inhibit lactation but would optimize the birth interval. Interestingly, it has been calculated that China could abandon the one-child-family policy without having to alter its population growth targets if the first birth was postponed until age 28, and there was at least a four-year birth interval^{35,36}. Well-spaced pregnancies would also reduce maternal mortality; according to one estimate, restricting the age of child-bearing to 20–39 could reduce maternal mortality by as much as 11% in developing countries³⁷. Prolonging the duration of lactational amenorrhoea may give additional maternal benefit because, if nothing else, it will preserve haemoglobin stores. Recent studies also suggest that, in the United States, breast feeding can nearly halve the risk of breast cancer relative to that of a parous woman who bottlefeeds her babies; the longer a woman breast-feeds, the greater the protection^{38,39}. Although this protective effect will be most pronounced in developed countries, with their high 7% lifetime incidence of breast cancer, it may also be of benefit to women in developing countries. If their duration of breast feeding was to decline, they might experience an increased incidence of breast cancer.

New disincentive

Recently, a new disincentive to breast feeding has emerged—the risk of transmitting the human immunodeficiency virus (HIV) from an infected mother to her child via breast milk. The World Health Organization held a consultation on this issue in June 1987 that reviewed all the scientific evidence and recommended that generally breast feeding should continue to be the feeding method of choice irrespective of the mother's HIV status, because the benefits to the baby outweighed the unproven risk of becoming infected from virus excreted in breast milk⁴⁰. However, the US Centers for Disease Control and the UK Department of Health and Social Security have recommended that HIV-positive women should not breastfeed^{41,42}. Such advice reflects the Western prejudice that artificial milks are innocent until proven guilty, whereas breast milk is guilty until proven innocent.

In the last analysis, it may not be possible to predict with accuracy the exact number of maternal, infant and childhood

deaths that could be saved by judicious combination of breast feeding and modern forms of contraception to achieve optimal birth spacing in excess of two years. We also lack the ideal contraceptive for this critical stage in a woman's reproductive life. But the potential benefits of breast feeding for the health of the mother and the child are so enormous that protection of breast feeding should always be uppermost when formulating public health policies for developed or developing countries—breast feeding and birth spacing save lives. □

We thank Peter Howie, D. B. Jelliffe, Chloe O'Gara, Roberto Rivera, Pramilla Senanayake, James Shelton, Jeffrey Spieler and Nancy Williamson for critically reviewing an earlier version of this article. Partial support for its preparation was provided by the United States Agency for International Development (USAID), although the views expressed by the authors do not necessarily reflect the views of USAID.

Shyam Thapa and Malcolm Potts are at Family Health International, Research Triangle Park, North Carolina 27709, USA. Roger V. Short is in the Department of Anatomy and Physiology, Monash University, Melbourne, Victoria 3168, Australia. Correspondence should be directed to Malcolm Potts.

1. Aristotle, *The Complete Works of Aristotle* Vol. 1 (ed. Barnes, J.) 1201 (Princeton Univ. Press, 1984).
2. Bongaarts, J. & Menken, J. in *Determinants of Fertility in Developing Countries* Vol. 1 (eds Bulatao, R. A. & Lee, R. D.) 27–60 (Academic Press, New York, 1983).
3. Bongaarts, J. & Potter, R. G. *Fertility, Biology and Behavior* (Academic Press, New York, 1983).
4. Short, R. V. *Sci. Am.* **250**, 35–41 (1984).
5. McNeilly, A. S. Breast feeding and Fertility. Seminar on Biomedical and Demographic Determinants of Human Reproduction (IUSSP, Liege, 1988).
6. Potts, M., Thapa, S. & Herbertson, M. A. *J. Biosoc. Sci. Suppl.* **9** (1985).
7. Gordon, K., Renfree, M. B., Short, R. V. & Clarke, I. J. *J. Reprod. Fert.* **79**, 397–408 (1987).
8. Peaker, M. J. *Physiol.* **301**, 415–428 (1980).
9. Loudon, A. S. I., McNeilly, A. S. & Milne, J. A. *Nature* **302**, 145–147 (1983).
10. Feachem, R. G. & Koblinsky, M. A. *Bull. WHO* **62**, 271–291 (1984).
11. Hanson, L. A. *Immunology Today* **3**, 168–172 (1982).
12. Walker, W. A. & Isselbacher, K. J. *New Engl. J. Med.* **297**, 767–773 (1977).
13. Minchin, M. *Birth* **4**, 25–34 (1987).
14. Bongaarts, J. in *Determinants of Fertility in Developing Countries*, Vol. 1 (eds Bulatao, R. A. and Lee, R. D.) 103–138 (Academic Press, New York, 1983).
15. Popkin, B. M., Bilsborrow, R. E. & Akin, J. S. *Science* **218**, 1088–1093 (1982).
16. World Health Organization, *Contemporary Patterns of Breastfeeding* (WHO, Geneva, 1981).
17. Ferry, B. & Smith, D. P. *World Fertility Survey Comparative Studies* No. 23 (International Statistical Institute, The Hague, 1983).
18. Notzon, F. *Pediatrics* **74**, 648–666 (1984).
19. Millman, S. *Int. Fam. Plann. Persp.* **12**, 91–95 (1986).
20. Page, H. J. & Lesthaeghe, R. (eds) *Child-Spacing in Tropical Africa: Traditions and Change* (Academic Press, New York, 1981).
21. Lesthaeghe, R. *Fertility and Its Proximate Determinants in Sub-Saharan Africa: The Record of the 1960s and 70s* (IUSSP, Liege, 1984).
22. Coale, A. J., Goldman, N. & Cho, L.-J. *Nuptiality and Fertility* (ed. Ruzicka, L. T.) 43–60 (IUSSP, Liege, 1979).
23. *World Fertility Survey Basic Documentation* No. 1 (International Statistical Institute, The Hague, 1975).
24. Berent, J., Jones, E. F. & Siddiqui, M. K. *World Fertility Survey Comparative Studies* No. 18 (International Statistical Institute, The Hague, 1982).
25. Hobcraft, J. N., McDonald, J. W. & Rutstein, S. O. *Popul. Stud.* **39**, 363–385 (1985).
26. Hobcraft, J. N., McDonald, J. W. & Rutstein, S. O. *Popul. Index* **49**, 585–618 (1983).
27. Pebley, A. R. & Millman, S. *Int. Fam. Plann. Persp.* **12**, 71–79 (1986).
28. Hobcraft, J. Background Paper, *Int. Conf. on Better Health for Women and Children through Family Planning*, (Nairobi, 1987).
29. Davanzo, J., Butz, W. P. & Habicht, J. P. *Popul. Stud.* **37**, 381–402 (1983).
30. Palloni, A. & Millman, S. *Popul. Stud.* **40**, 215–236 (1986).
31. Fortney, J. A. & Higgins, J. E. *Public Health* **98**, 73–83 (1984).
32. Bongaarts, J. *Pop. Devel. Rev.* **13**, 323–334 (1987).
33. Trussell, J. *Pop. Devel. Rev.* (in the press).
34. Cleland, J. & Rutstein, S. *Int. Fam. Plann. Persp.* **12**, 83–90 (1986).
35. Bongaarts, J. & Greenhalgh, S. *Pop. Devel. Rev.* **11**, 585–618 (1985).
36. Trussell, J. *Int. Fam. Plann. Persp.* **12**, 80–82 (1986).
37. Trussell, J. & Pebley, A. R. *Stud. Fam. Plann.* **15**, 267–280 (1984).
38. McTiernan, A. & Thomas, D. B. *Am. J. Epidemiol.* **124**, 353–358 (1986).
39. Kvale, G. & Heuch, I. *J. Epidemiol. Comm. Health* **42**, 30–37 (1987).
40. World Health Organization *Breastfeeding/Breast Milk and Human Immunodeficiency Virus* (WHO, Geneva, 1987).
41. *Centers for Disease Control MMWR* **34**, 721–732 (1985).
42. Present day practice in infant feeding *Report on Health and Social Subjects* **32**, 18 (HMSO, London, 1988).
43. Casterline, J. B., Singh, S., Cleland, J. & Ashurst, H. *World Fertility Survey Comparative Studies* No. 39 (International Statistical Institute, The Hague, 1984).
44. Cleland, J. & Hobcraft, J. (eds) *Reproductive Change in Developing Countries* (Oxford University Press, 1985).
45. Fildes, V. *Breast, Bottles and Babies: A History of Infant Feeding* (Edinburgh University Press, 1987).

How eukaryotic transcriptional activators work

Mark Ptashne

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

A specific protein, bound to DNA, can activate transcription of a wide array of genes in many eukaryotes. Further analysis suggests a general outline for how eukaryotic transcriptional activators function and are controlled.

Two years ago we had at hand, arguably, the outline of an answer to the following question: how can a protein, bound to a site hundreds or even thousands of base pairs from a gene, activate transcription? According to our argument, an activator binds DNA and interacts with (touches) another protein bound near the transcription start, the intervening DNA looping out to accommodate the reaction^{1,2}. Implicit in this description is the notion that the activator protein must have two surfaces: one, the DNA-binding region, positions the protein on the DNA so that the other, the activating region, can interact with the target protein to promote initiation of transcription. To my knowledge no experiments reported since then strongly contradict this scheme, and a number of results support it, if indirectly.

In this article I trace some of the background to these ideas, and then review recent experiments that further clarify our understanding of how eukaryotic transcriptional activators work. Experiments performed primarily with yeast activators suggest the following generalizations: (1) Activators are modular proteins whose activating regions are readily interchangeable; so are their DNA-binding regions. (2) Whereas DNA-binding surfaces are specific structures, each with a highly defined specificity for its cognate DNA-binding site, activating regions are less precisely defined structures characterized by an excess of acidic residues. Some aspect of structure—perhaps an α -helix—is essential for formation of an activating region. (3) An activator that works in yeast will also work in mammalian, plant and insect cells, provided the test gene bears in its vicinity a site recognized by the DNA-binding surface. (4) Any two of these activators can work cooperatively to stimulate transcription. (5) Activators with stronger activating regions more readily work at greater distances on the DNA. (6) Activators with stronger activating regions are controlled so that these regions are exposed in the nucleus in their fully functional form only when they are required to work.

Some of these generalizations are rather more speculative than others, but all are consonant with a variety of observations concerning regulation of transcription of eukaryotic genes. Moreover, they suggest a coherent picture of how eukaryotic activators are constructed, how they turn on transcription, and how they are controlled.

Transcriptional activators

The idea that a transcriptional activator bears two distinct surfaces, one to bind DNA and the other to touch another protein, arose from analysis of gene activation by the λ phage repressor. As illustrated in Fig. 1, λ repressor binds adjacent to RNA polymerase and helps it to initiate transcription of the repressor gene. The important conceptual point was established by the isolation of mutant repressors that bind DNA but fail to activate^{3,4}. These mutants bear changes on the surface of repressor that, as suggested by models of repressor and polymerase bound to DNA, most closely approaches RNA polymerase. The activation function requires the DNA-binding function: mutants

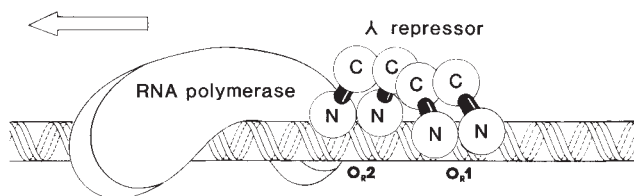


Fig. 1 λ repressor working as an activator. The repressor dimer bound to site O_{R1} helps another bind to O_{R2} , and the latter contacts RNA polymerase to stimulate transcription. Both the DNA-binding and activating surfaces of λ repressor are found on its amino terminal domain. The carboxyl domain is responsible for the most important contacts holding each dimer together and also for the cooperative interaction between dimers. Two repressor dimers also bind cooperatively to a DNA molecule bearing O_{R1} moved further upstream an integral number of helical turns from O_{R2} , the intervening DNA looping out to accommodate the reaction^{74,75}. But in that case the repressor at O_{R2} cannot activate transcription, possibly because interaction with the distally bound repressor constrains the repressor at O_{R2} so that it cannot appropriately contact RNA polymerase⁷⁶.

unable to bind DNA cannot activate because the protein must be positioned on DNA near polymerase for the required protein-protein interaction. The activating surface of λ repressor consists in part of an amphipathic α -helix bearing solvent-exposed residues that are predominantly negatively charged (Asp and Glu). Mutations that damage the activation function decrease the negative charge, and a second-site revertant of one such mutant restores that negative charge. Comparison of the activating regions of λ with those of the 434 and P22 repressors reveals solvent-exposed acidic residues in all of them. Insertion of these residues from λ repressor into the appropriate surface of λ Cro—a protein that binds to the same site on DNA as λ repressor but which cannot activate—confers on that protein the ability to activate transcription⁵.

The DNA-binding and activating surfaces of λ repressor lie on the same domain of the protein. In contrast, these two functions are found on different domains in eukaryotic activators, a conclusion suggested by a series of experiments demonstrating that the two functions are readily separable. In the first of these 'domain swap' experiments, the DNA-binding domain of a yeast transcriptional activator (GAL4) was replaced with that of a bacterial repressor (LexA). The hybrid protein synthesized in yeast activates transcription of a yeast gene bearing an upstream LexA operator⁶, whereas an amino-terminal fragment of GAL4 containing its DNA-binding region, but not its activating region, binds DNA and fails to activate⁷. These experiments argue that for a eukaryotic activator, as for λ

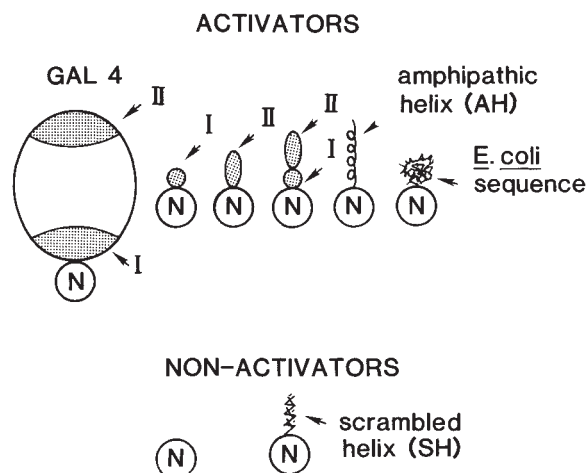


Fig. 2 GAL4-related proteins. The circled *N* designates GAL4 (1–147), a fragment of GAL4 that binds DNA. The largest structure is intact GAL4 with its two acidic activating regions; regions I and II include, respectively, amino acids 148–238 (net charge –7) and residues 768–881 (net charge –9). The three adjacent activators carry GAL4's activating region I or region II, or both. AH is a 15-amino-acid sequence designed so that it can form an amphipathic helix bearing negative charges along one face, and the *E. coli*-derived sequences are acidic activating regions found encoded in the *E. coli* genome. The two non-activators include GAL4 (1–147) as well as GAL4 (1–147) carrying the 15 amino acids found in AH but in scrambled order (SH). The ordinary site of action of GAL4 in yeast is called the galactose upstream activating region (UAS_G), which includes four closely spaced 17-base-pair sites^{77,78}; each of these sites is partially twofold rotationally symmetric and binds a dimer of GAL4 (M. F. Carey and M.P., unpublished). A single 17-base-pair site, synthesized as a near consensus of the naturally occurring sites, will also mediate activation by GAL4 derivatives⁷⁷.

repressor, DNA binding is a 'neutral' function which brings the activating surface of the protein to the vicinity of a gene. LexA-GAL4 activates transcription from a LexA operator positioned at many locations within several hundred base pairs upstream of the transcription start site, just as wild-type GAL4 activates transcription from its ordinary DNA-binding sites inserted at various positions upstream of any of several yeast genes^{8,9}.

A modified form of the domain-swap experiment, involving instead 'domain addition', showed that the activating region can function in either orientation in relation to the DNA-binding region. In this experiment, the LexA DNA-binding domain was attached to the amino terminus of the intact yeast regulator GCN4 (ref. 6). In the hybrid protein, the activating region of GCN4 lies between the two DNA-binding domains—that of GCN4 itself (at the carboxyl terminus)^{10,11} and that of the LexA domain—and the hybrid activates transcription when bound either to a LexA (ref. 6) or to a GCN4-binding site (ref. 11). Subsequent experiments suggest that this kind of flexibility is found in other eukaryotic activators as well; for example, in one case it has been shown that the activator consists of domains whose positions in the molecule are interchangeable¹².

Nature of activating regions

Deletion analysis of the LexA-GCN4 protein described above¹¹, and then of GAL4 (ref. 13), defined activating regions as stretches of amino acids bearing a significant net negative charge. In GCN4, a protein of 281 amino acids, the activation region extends over some 60 amino acids with a net charge of –16, whereas GAL4, which is much larger (881 amino acids), contains two negatively charged regions of ~100 amino acids each, either of which activates transcription when attached to the DNA-binding domain. When both activating regions of GAL4 are fused to a fragment GAL4 (1–147) that contains the DNA-binding domain, activation is nearly as efficient as with intact GAL4; activation by GCN4 is roughly proportional to the fraction of the activating region attached to its DNA-binding region¹⁴.

The acidic character of activating regions is evidently not coincidental. For example, starting with the small protein GAL4 (1–238) (which contains the DNA-binding region plus activating region I, see Fig. 2), point mutants are readily isolated that have increased activation functions, the typical single mutation having about a threefold effect¹⁵. With one exception that I discuss below, mutants that increase activation bear amino acid changes that increase the negative charge; in several cases the mutations replace positively charged residues with neutral ones. Individually isolated mutations are approximately additive: sequential addition, by recombination *in vitro*, of four mutations

to the GAL4 (1–238) gene increases stepwise the activity of the encoded protein. The final product, which has four more negative charges than the parent, can activate transcription ninefold more efficiently than the parent, and 80% as efficiently as intact GAL4 (ref. 16).

This combined evidence suggests that no very elaborate structure is required to display the negative charges on an activating region. This conclusion was reinforced by the demonstration that sequences that can function as activating regions are readily found encoded in random bits of the *Escherichia coli* genome¹⁷. One of these, designated B42, can be fused to GAL (1–147) to make an activator which functions nearly as efficiently as intact GAL4, and this sequence will also work when transferred to the LexA DNA-binding domain. The activity of LexA-B42 shows that a yeast activator can be constructed entirely from sequences derived from *E. coli*.

Although I have emphasized the importance of negative charge, we know that charge is not the sole characteristic of an activating region. This conclusion was first suggested by the fact that mutants of GAL4 with decreased activation function do not always bear fewer negative charges than the parent¹⁵. The following observations and experiments suggest that an α -helix might be an essential element of an activating region.

Inspection of activating region sequences, including those encoded by bits of the *E. coli* genome, suggests that many of them could, in principle, form amphipathic α -helices¹⁸. These helices would bear negatively charged residues along one surface and hydrophobic residues along another, a feature shared by the activating region of the λ phage repressor. This consideration prompted the construction of two oligonucleotides, each encoding a 15-amino-acid peptide and each attached to DNA encoding GAL4 (1–147). One peptide (AH) was designed so that, should it form an α -helix, that structure would be amphipathic, bearing negatively charged residues along one surface and hydrophobic residues along another. The second peptide (SH) comprised identical amino acids, but in scrambled order. The construct bearing the first peptide activated transcription in yeast, whereas that bearing the second peptide did not. Deletion analysis of GCN4 also suggests that an α -helix could be an important component of an activating region¹⁴.

We now return to the anomalous mutant of GAL4 (1–238), in which increased activation is not associated with an increase in the acidity of the primary sequence. Preliminary experiments suggest that this mutant, in which an Ile is replaced by a Thr, is phosphorylated (I. Sadowsky, G. Gill and M.P., unpublished results). Phosphorylation may increase the potency of an activating region, perhaps simply by contributing negative charge¹⁹, and could thereby provide a possible means of controlling activators in response to environmental signals^{20–22}.

A group of hormone-responsive activators, including the human oestrogen and the rat and human glucocorticoid receptors, has recently been found to fit the activator paradigm we have been describing. That is, each of these proteins contains one or more activating regions that are separable from the DNA-binding domain. In the rat glucocorticoid receptor, for example, the DNA-binding region has been replaced by the DNA-binding domain of LexA, and the hybrid can bind to LexA operators and activate gene expression in mammalian cells²³. Deletion analysis of these LexA-receptor hybrids reveals an acidic activating region in the amino-terminal portion of the molecule. Similar hybrids between a GAL4 DNA-binding fragment and the human glucocorticoid receptor bind to GAL4-binding sites and activate transcription in mammalian cells. These molecules have, in addition to an activating region in the amino-terminal portion of the molecule, a second activating region in the carboxyl part^{24,25}; the replacement of one of these activating regions with two copies of the short amphipathic α -helix described above produces a molecule that retains nearly full activity²⁵. Further analysis will be facilitated by the finding that the glucocorticoid and oestrogen receptors activate transcription in yeast^{26,27}.

'Domain-addition' experiments have been adduced to argue that the products of the oncogenes *fos* and *myc* are transcriptional activators. Thus, fusion of either protein to the DNA-binding domain of LexA produces a hybrid that binds to the LexA operator and activates transcription in yeast. As pointed out by the authors¹⁹, the result is only suggestive; it will be interesting to see whether mutants deficient in the activating function assayed in yeast are similarly deficient in the transformation function when assayed in mammalian cells.

Heterologous gene activation

Various members of the GAL4 activating family shown in Fig. 2 activate transcription in mammalian, insect and plant cells as well as in yeast. A common strategy was used to demonstrate activation in these various organisms. Thus in each case two DNA molecules were introduced into cells: one (the 'effector'), expresses GAL4 or one of its derivatives, and the other (the 'reporter') carries a gene modified so that it bears GAL4-binding sites nearby; in most cases the reporter gene was also fused to a gene whose product is readily assayed by enzymatic activity. In all of the experiments that follow, the reporter showed no activity in the absence of GAL4-binding sites.

Drosophila larvae. This demonstration of GAL4's function exploits the pattern of tissue-specific activity of the *Drosophila Adh* promoter²⁸. The effector carries the *GAL4* gene fused to the *Adh* promoter, and the reporter the *lacZ* gene of *E. coli* fused to a *Drosophila* promoter. In larvae bearing both constructs, only tissues in which the *Adh* promoter is known to be active express β -galactosidase. The level of expression elicited by GAL4 is at least as great as that seen when the *lacZ* gene is fused directly to the *ADH* promoter.

Plant cells. Three GAL4 derivatives that activate in yeast cells also activate transcription from a plant viral promoter in tobacco cells in culture²⁹. In this experiment the GAL4 derivatives are expressed from the nopaline synthase promoter, and the reporter carries a modified form of the 35S cauliflower mosaic virus promoter bearing multiple GAL4-binding sites.

Mammalian cells. GAL4 and various derivatives activate an array of mammalian promoters when assayed in tissue culture. Examples include the promoter of mouse mammary tumour virus and promoters of the *tk* and β -globin genes^{30,31}. Each of the proteins of Fig. 2 defined as activators in yeast also works as an activator in mammalian cells, whereas the two defined as non-activators in Fig. 2 do not^{30,31} (I. Sadowski and M.P., unpublished observations). Moreover the quadruple mutant of GAL4 (1-238), which bears increased negative charge and activates transcription in yeast more efficiently than the parent, also activates more efficiently in mammalian cells.

In most of the experiments described in this section so far, the GAL4-binding sites were placed close to the TATA box sequence found near the transcriptional start of many eukaryotic genes. But, as already noted, genes of higher eukaryotes are often activated by proteins that bind to sequences called enhancers, which in some cases are positioned many hundreds of base pairs upstream or downstream of the gene³². In yeast GAL4 ordinarily acts at an intermediate distance (~250 base pairs)⁹. Therefore the question arises of whether activators of the sort we have been describing can confer the properties of enhancers on their binding sites. The answer is yes^{31,33}, and the following additional consideration and result reinforces an idea implicit in our discussion of activating regions. It is reasonable to assume that activating regions differ, because of charge and structure, in the avidities with which they interact with a target protein, and moreover that an activating region with high affinity for the target would be more effective at activating transcription at a distance. Indeed we find that a particularly strong activating region attached to the GAL4 DNA-binding domain will efficiently activate transcription from sites over 1,000 base pairs from the start point of transcription.

Such a particularly potent activating region is found on the herpes simplex virus protein VP16 (refs 35-39). This protein is a component of the virion that activates transcription of immediate early genes in virally infected cells. The remarkable efficiency of this activating region is illustrated by the properties of a hybrid protein consisting of the highly acidic amino-acid sequence found at the carboxyl terminus of VP16 and fused to a DNA-binding fragment of GAL4 (ref. 33). For example, with GAL4-binding sites located close to TATA, the hybrid activated transcription in mammalian cells at least 100-fold more efficiently than did GAL4 bound to the same sites; when the sites were positioned about 2,000 base pairs downstream, the VP16 hybrid activated efficiently, whereas GAL4 did not. Another very strong activating sequence is found in a second viral protein, E1A, a transcriptional activator encoded by adenovirus. Thus, fusion of the appropriate 100-amino-acid segment of E1A to the GAL4 DNA-binding fragment produces an activator that, when bound to GAL4-binding sites on DNA, activates transcription efficiently even when those sites are over 1,000 base pairs from the gene. (J. W. Lillie and M. R. Green, personal communication). So it is clear that activators bearing the same DNA-binding domain, but with different activating regions, can differ markedly in their abilities to activate. At the end of the next section I suggest a mechanism whereby activators with weaker activating regions can activate at a distance.

I have so far stressed the success of our ideas in explaining gene activation, particularly activation at a distance. But it would be misleading to suggest that we understand all aspects of the problem even as described to this point. For example, an array of auxiliary transcription factors have been described that bind near the transcription start site of mammalian genes⁴⁰, and we do not know what role these factors might have in activation. The simplest view is that these proteins are further examples of acidic activators with weak activating regions which perform the same function as those we have been describing. An alternative possibility is that these proteins belong to a different class and perform some special function in helping distally bound activators to work—a function that is not required if the distal activators are moved close to the gene⁴⁰. To take another example, the GAL4-E1A hybrid discussed above differs from all the other activators discussed here in that, for reasons that we do not understand, the hybrid activates transcription, albeit more weakly, in the absence of GAL4-binding sites. In this regard, the hybrid resembles E1A itself, which can activate in the absence of known specific DNA-binding sites⁴¹. Despite these uncertainties, the argument in the following section suggests that it is not coincidental that the two strongest activating regions found so far are on viral proteins.

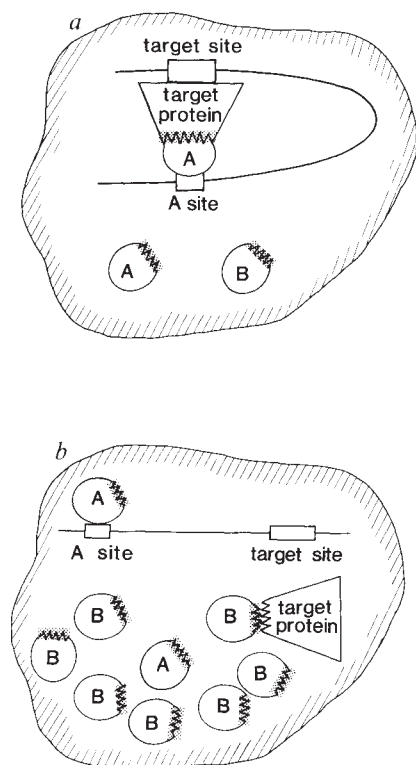


Fig. 3 Squelching. In the cell in *a*, the activators A and B are present in ordinary concentrations. The target protein is brought to the DNA by interaction with A bound to its site (a protein-protein interaction), and by interaction with the target site (a protein-DNA interaction). In the cell in *b* the activator B is present in great excess and binds to the target protein off the DNA; because the activating regions of both A and B see the same site on the target protein, interaction with B prevents that protein from interacting with A, and the gene is off. An alternative explanation for squelching—the formation of mixed oligomers—seems unlikely but is not excluded (see ref. 16 for discussion).

Squelching

I began this review by assuming that DNA binding brings to the DNA, and hence to the vicinity of the gene, an activating region which in turn interacts with some component of the transcriptional machinery. The latter interaction might bring this target protein to the DNA, helping it to bind to its site and to participate in formation of an active transcription complex. A corollary of this view is that an activating region might also interact (perhaps more weakly) with the target protein when both activator and target are off the DNA. Figure 3 illustrates how such an interaction could have an inhibitory effect: assuming that the target protein is present in limiting amounts, interaction between it and the activator off the DNA could make the target protein unavailable for transcription. At high concentrations, activators could depress transcription from promoters lacking the activator binding sites, and at still higher concentrations activation of promoters bearing the binding sites would be inhibited as well. For any given concentration of activator, this inhibitory effect would be stronger the more potent the corresponding activating region. Experiments in yeast and mammalian cells are consistent with these expectations, and we call the phenomenon 'squelching' (squelch: to put down, suppress or silence⁴²).

In yeast, overproduction of intact GAL4 inhibits transcription of a typical gene lacking GAL4-binding sites, the *CYC1* gene

for example¹⁶. Experiments with the various GAL4-derived activators show a good correlation between inhibitory and activating strengths. Squelching does not require DNA binding: GAL4 (79–881), which lacks the DNA-binding determinants, inhibits as effectively as intact GAL4.

There is a further interesting correlation between activation and squelching. Struhl⁴³ has identified a transcript in yeast, expressed at a low level, that is not inducible by acidic activators. According to the line of reasoning outlined above, we would expect that formation of this transcript does not require the target protein seen by acidic activators and therefore would not be inhibited by high levels of GAL4, and experiments confirm this prediction¹⁶.

Squelching is also observed in mammalian cells transiently expressing the potent activator GAL4-VP16 (ref. 33). This effect parallels the observations of Triezenberg *et al.*³⁵, who showed that native VP16 depresses expression of a reference gene as it activates expression of a test gene bearing the appropriate sites; deletion variants of VP16 show that stronger activators are also stronger squelchers.

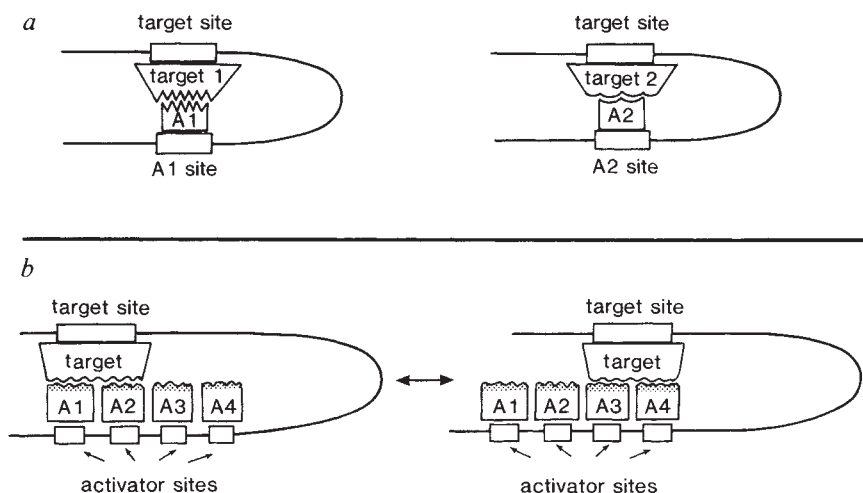
These ideas and results stress the importance of regulating the concentrations of activators in cells, and of ensuring that the strongest activators are exposed only transiently, as for example, in viral infection. Before returning to a more complete discussion of how activators are regulated, I reconsider the apparent requirement for very strong activators to effect action at a distance. Expression of such very strong activators, we have seen, is unavoidably associated with the risk of squelching. But we know that natural activators, present in fully functional form in cells, can activate transcription at great distances. How might these cellular activators work without the risk of squelching? The answer, suggested explicitly or implicitly by others^{44–48}, is that activation in these cases involves the combination of many weak interactions, rather than one or a very few strong ones. That is, multiple weak activators might cooperate to interact efficiently with the target when bound to DNA, although none taken separately would have activating regions strong enough to interact significantly with the target protein off the DNA. In the following section I describe examples of promiscuous cooperativity between acidic activators using, in part, derivatives of GAL4, a molecule which itself works cooperatively in its ordinary action in yeast⁴⁹.

Promiscuous cooperativity

GAL4 and some of its derivatives work cooperatively (synergistically) with a variety of transcriptional activators from other organisms^{30,31,50}. A striking example comes from an experiment performed *in vitro* (see next section), in which a GAL4 derivative was found to work cooperatively with the mammalian transcriptional activators ATF (activating transcription factor)⁵¹ and USF (upstream stimulating factor)^{52,53}. Thus, whereas either the GAL4 derivative or one of the mammalian transcription factors used separately stimulated transcription perhaps two to fourfold, working together they stimulated transcription more than 50-fold⁵⁰. What is evidently the same phenomenon has also been observed *in vivo* in mammalian cells, in which GAL4 and the glucocorticoid receptor were shown to be able to activate transcription synergistically³⁰, and in yeast cells, where GAL4 derivatives can work cooperatively with LexA-B42, the activator consisting entirely of *E. coli*-derived sequences described earlier. (H. J. Himmelfarb and M.P., unpublished).

What feature might these disparate molecules (GAL4 derivatives on the one hand and mammalian or *E. coli*-derived activators on the other) have in common that enables them to work cooperatively? One of the possibilities I consider in Fig. 4 and its legend is that any two eukaryotic activators can work synergistically not because they touch each other, but rather because they simultaneously touch a third (target) protein. That mechanism would differ in two ways from the well understood case involving λ repressors working cooperatively to simulate

Fig. 4 Two ideas for how eukaryotic genes might be activated. *a*, Two active genes, one activated by A1, the other by A2. The activating regions of A1 and A2 see different targets, for example, different forms of the TATA-binding protein. Each activator has a precisely defined and different activating region that docks with a unique surface on its cognate target. The DNA between the activator and the target is looped out. *b*, One active gene. The target protein samples the acidic activating regions of four DNA-bound activators, perhaps two at a time. That interaction brings the target protein (e.g., TFIID) to its site, and the continual sampling keeps it there and holds it in a conformation so that RNA polymerase can attach and begin transcription. Any two activators, bound to DNA, that simultaneously touch the target protein will work cooperatively⁷⁹, and additional DNA-bound activators will add their effects more nearly linearly. The description implies that the activators are firmly bound to DNA before interaction with the target protein. But we could picture the various protein-protein and protein-DNA interactions as concerted, in which case the synergistic effect of the activators could be reflected in cooperative binding of the activators to their sites on DNA. That effect would most readily be observed with DNA sites that bind the activators weakly. The reader will appreciate that our picture here is highly conjectural and that other models are easily imaginable. For example, it is conceivable that the acidic activating regions on the various activators interact directly with each other, perhaps by apposition of hydrophobic surfaces of amphipathic α -helices. In that case cooperative DNA binding of the activators would not necessarily involve simultaneous interaction with the target. Models similar to that of Fig. 4*b* have been suggested by others^{44,45}.



transcription as described in Fig. 1. First, in the λ case, only one repressor molecule contacts the target protein (in this case RNA polymerase), and second, cooperative repressor binding depends on a specific protein-protein interaction between identical proteins that does not occur between different repressors (such as the λ and 434 repressors). Cooperativity in the λ case contributes to the formation of a sensitive switch that responds dramatically to a small change in repressor concentration², and presumably cooperativity, whatever the mechanisms, can have an analogous role in eukaryotes.

What do activating regions see?

The activating regions we have been discussing presumably interact with a component common to many transcription complexes in many organisms (see next section). The polymerase itself, one obvious possibility⁵⁴, is a multicomponent enzyme, and various accessory proteins aid in the formation of transcription complexes. One of these is TFIID, a protein described by Roeder and his colleagues, which binds to the TATA sequence in front of many genes and is required for transcription *in vitro*⁵². Experiments with partially purified components suggest that mammalian transcriptional factors interact with TFIID. For example, when assayed on the appropriate promoter, DNA-bound ATF changes the conformation of TFIID bound at the TATA region, extending the region on DNA covered by the protein, and this reaction may suffice to stimulate transcription⁵⁵. TFIID is thus another possible target for the action of eukaryotic activators, and experiments with GAL4 (1-147) + AH support this possibility. This GAL4 derivative, synthesized in and purified from *E. coli*, will stimulate transcription of a gene in a mammalian cell extract⁵⁰ and can be shown to have an effect on the conformation of DNA-bound TFIID which is identical to that of the mammalian transcription factors⁵⁶. The TATA-binding protein is interchangeable between yeast and mammals^{57,58}, a finding that could account for the activation of mammalian genes by yeast activators and *vice versa*. If the common target seen by acidic activating regions is in fact the TATA-binding protein, similar TATA-binding proteins should

also be present in many other organisms. More precisely, the various TATA-binding proteins should have at least two surfaces in common: one to see the TATA sequence and the other to interact with the acidic activating sequence.

A scheme for activation

One synthesis of the findings I have described is presented in Fig. 4*b*, the salient features of which are emphasized by comparison with the model in Fig. 4*a*. In the model of Fig. 4*a*, each activating region of an activator is a highly specific structure that interacts with a unique target protein, and an active gene is found in a precisely defined protein-DNA complex. According to that model, a given regulator placed on DNA in the vicinity of a gene would activate that gene only if it belonged to the subset of genes to which the unique target protein could also bind.

But many of the findings reviewed here, as well as other widely appreciated features of gene regulation in eukaryotes, argue that for many genes the model shown in Fig. 4*a* cannot apply. For example, we have known for some time that the typical gene of a higher eukaryote bears in its vicinity an array of distinct binding sites for regulatory proteins⁴⁰. Attaching these sites to a typical second gene in place of its ordinary regulatory elements does not inactivate that gene, but rather confers on it the pattern of expression characteristic of the first gene^{8,40}. Different DNA control elements can often be mixed together, or some control elements deleted or duplicated, with the result that new (and often unpredicted) patterns of tissue-specific expression are conferred⁵⁹. One surmises that, contrary to a prediction from the model in Fig. 4*a*, the total number of activators is small compared with the number of regulated genes in a higher eukaryote. Very often common elements are used in different combinations to confer the many patterns of tissue-specific gene expression. This flexibility would be ruled out by the requirement for a unique interaction between a given regulator and its target, as in the model of Fig. 4*a*.

The model shown in Fig. 4*b* lacks the disadvantages of the model in Fig. 4*a*, but it suggests that the active gene is part of

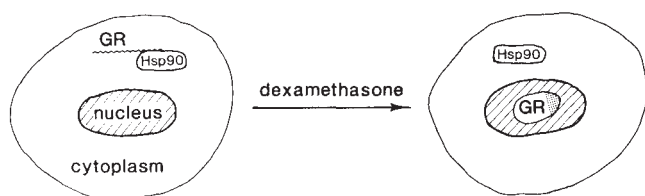


Fig. 5 Control of the glucocorticoid receptor (GR). The protein HSP90 binds to the receptor and holds it, perhaps partially denatured, outside the nucleus. Hormone dissociates the complex and allows the receptor to move to the nucleus and activate transcription. The control module of the receptor which binds HSP90 can be attached to a heterologous molecule, conferring hormone inducibility to that protein⁶². The oestrogen receptor is also believed to be inhibited by HSP90⁸⁰. One of the important inhibiting effects of HSP90 may be to prevent receptor monomers from forming dimers, the activating species⁸¹ (see ref. 24 for further discussion).



Fig. 6 Control of GAL4. The inhibitor GAL80 covers the activating regions, but not the DNA-binding region, of GAL4; growth in galactose dissociates the complex and allows GAL4 to activate transcription. GAL4 is shown binding as a monomer, but, as pointed out in the legend to Fig. 2, GAL4 binds as a dimer to each site.

a less precisely defined structure. We might invoke the analogy of ion-exchange chromatography, imagining that the target protein continuously samples, perhaps two or more at a time, the exposed activating regions of the DNA-bound activators. Depending on the number and strength of these interactions, the target protein remains bound and transcription ensues. The model in Fig. 4b predicts that a given activator can activate any of a wide array of genes; that all activating regions present some common feature for sampling; that there are no strong requirements for the structure of activating proteins other than that each comprises a DNA-binding and an activating surface; and that any two activators can work synergistically (see Fig. 4 legend). In the course of this review we have encountered results consistent with each of these predictions.

In the context of our overall picture, it is perhaps not surprising that activating regions are characterized in the first instance as bearing excess negative charge. It would seem reasonable that a system in which many different proteins (activators) must interact with a common target would exploit electrostatics. Other interactions—hydrogen bonds for example—are more geometrically constrained and work over shorter distances, and it might be more difficult to build in the required redundancy using such interactions.

Other interactions of regulatory proteins

So far I have emphasized the interaction of activating regions with some common target to activate transcription. But the activity of eukaryotic transcriptional activators is often controlled by interaction with proteins other than the common target. Here I consider interactions of this kind involving some of the proteins we have already discussed, and I then suggest a generalization concerning the control of activators that draws on various themes touched on in the course of this article.

Each steroid hormone receptor activates gene expression only if its cognate steroid is present in the medium. The glucocorticoid receptor, for example, is believed to bind to the protein HSP90

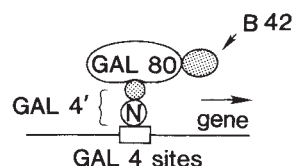


Fig. 7 Turning GAL80 into an activator. The acidic activating sequence B42 is added to a non-essential part of GAL80. The hybrid binds to a DNA-bound derivative of GAL4 that itself activates only weakly, and activates transcription. The carboxyl 30 amino acids of GAL4, a component of activating region II, suffice for recognition by GAL80⁶⁴, and the GAL4 derivative used in this experiment (GAL4') consists of GAL4 (1-147) fused to these 30 amino acids.

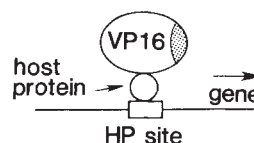


Fig. 8 Activation by VP16. This viral protein does not itself recognize DNA. Rather, it is brought to the DNA by interaction with a DNA-bound cellular protein, in such a way that its acidic activating region is free to activate transcription. Note that the host protein, although essential for activation, need not itself be an activator in the sense we have been using the term in this article. References 82 and 83 describe two examples of protein complexes that bind to DNA for which we have a more complete understanding of mechanism; one of these complexes sees an asymmetric site⁸³.

to form a complex that is held in the cytoplasm. According to this idea, hormone releases HSP90 and thereby allows the receptor to move to the nucleus, bind DNA, and activate gene expression^{12,60-63} (see Fig. 5).

GAL4 turns on the genes required for galactose metabolism in yeast only if galactose is present in the medium. This control involves an inhibitor (GAL80) that binds to GAL4 and covers its activating regions; galactose, or a metabolic derivative, releases GAL80 from GAL4⁶⁴⁻⁶⁶ (see Fig. 6). The GAL4-GAL80 complex binds to the GAL4-binding sites on DNA; one demonstration of this fact is that GAL80 can be converted from an inhibitor to an activator by inserting into it the activating sequence B42 discussed previously⁶⁷. To work as an activator, GAL80-B42 must be brought to the DNA by interaction with a DNA-bound GAL4 (Fig. 7). In this regard GAL80-B42 mimics VP16, which is not itself a DNA-binding protein; rather, VP16 recognizes proteins bound at either of two DNA sites located upstream of the gene³⁴⁻³⁹ (see Fig. 8).

Control of activators

I have argued that in eukaryotes transcriptional activators recognize a common target; that activators with strong activating regions readily work at a distance; and that high expression of a strong activating region is unavoidably detrimental to cell viability because of squelching. Taken together, these propositions suggest a further idea: simultaneous exposure of strong, or even moderately strong, activating regions found on different eukaryotic activators might be harmful. This suggests, in turn, that these activators should be controlled so that each exposes its activating regions only when it is required to work. Indeed this seems to be the case from the examples we have considered: in the absence of galactose, GAL4's activating region is covered by GAL80; in the absence of hormone, steroid receptors are held in an inactive complex with HSP90; and VP16 and E1A are made only transiently. It is also likely, as we have noted, that in many cases an exposed eukaryotic activating region will

function only if phosphorylated, and that phosphorylation can be subject to physiological control (for other examples, see refs 22 and 68).

This situation should be contrasted with that observed for a class of activators in *E. coli*. Most well characterized activators of this class (such as λ repressor, CAP, AraC) are (or can be) present in the cell's nucleoid body with their activating regions exposed even when they are not working. But these activating regions are presumably very weak²: they activate transcription only when the protein is positioned very near RNA polymerase. Control is manifested at the level of DNA binding; for example, CAP binds DNA only when complexed with cyclic AMP. A tentative generalization suggests itself: stronger activating regions will not be continuously exposed in the nucleus (or nucleoid body), whereas others generally will be.

That we are not dealing here merely with a distinction between bacteria and eukaryotes is suggested by, on the one hand, the presence in eukaryotic nuclei of fully functional weak activators and, on the other hand, by the properties of NR1, an activator of the glutamine synthetase gene of *E. coli* that works efficiently at a distance⁶⁹. Phosphorylation of this bacterial activator, a programmed response to nitrogen starvation, controls the activator primarily (but not entirely)⁷⁰ by potentiating its activation function^{71,72}.

E. coli—a puzzle

Preliminary experiments indicate that *E. coli* is at least one

organism in which our generalized activators do not work (S. D. Liang and M.P., unpublished results). Perhaps those with sensitive noses should detect a waft of irony here: recall the similarity between the activating surface of λ repressor and the putative α -helix that can function as a eukaryotic activating region. But in *E. coli*, even a natural bacterial activator may be able to turn on only a subset of genes. That is, different groups of *E. coli* genes are transcribed by different forms of RNA polymerase distinguished by their σ subunits. Five different σ subunits have been identified, each of which directs RNA polymerase to a specific subset of promoter sequences⁷³, and the evidence we have suggests that an activator that can turn on promoters using one type of σ factor does not turn on promoters that use other types⁶⁹. It is tempting to speculate that one of these σ factors is similar to the characterized TATA-binding protein, and that an as-yet-untested σ will render its cognate bacterial promoter activatable by our generalized eukaryotic regulators. Whether or not this turns out to be the case, the situation in *E. coli* should alert us to the possibility that there will be, despite the generalities described here, classes of RNA polymerase II-transcribed genes that respond to activators different from the ones I have described.

I am grateful for numerous helpful conversations I enjoyed during the course of writing this review. I also thank R. Axel, M. Botchan, P. Chambon, E. Giniger, L. Gold, A. Johnson, S. McKnight, H. Nash, H. Pelham, K. Yamamoto and students and colleagues at Harvard for comments on the manuscript.

1. Ptashne, M. *Nature* **322**, 697-701 (1986).
2. Ptashne, M. *A Genetic Switch* (Cell Press, Palo Alto, California & Blackwell Publishing, 1986).
3. Guarente, L., Nye, J. S., Hochschild, A. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2236-2239 (1982).
4. Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32**, 319-325 (1983).
5. Bushman, F. D. & Ptashne, M. *Cell* **54**, 191-197 (1988).
6. Brent, R. & Ptashne, M. *Cell* **43**, 729-736 (1985).
7. Keegan, L., Gill, G. & Ptashne, M. *Science* **231**, 699-704 (1986).
8. Guarente, L., Yocum, R. & Gifford, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7410-7414 (1982).
9. West, R. W. Jr, Yocum, R. & Ptashne, M. *Molec. cell. Biol.* **4**, 2467-2478 (1984).
10. Hope, I. A. & Struhl, K. *Cell* **43**, 177-188 (1985).
11. Hope, I. A. & Struhl, K. *Cell* **46**, 885-894 (1986).
12. Picard, D., Salser, S. J. & Yamamoto, K. *Cell* (in the press).
13. Ma, J. & Ptashne, M. *Cell* **48**, 847-853 (1987).
14. Hope, I. A., Mahadevan, S. & Struhl, K. *Nature* **333**, 635-640 (1988).
15. Gill, G. & Ptashne, M. *Cell* **51**, 121-126 (1987).
16. Gill, G. & Ptashne, M. *Nature* **334**, 721-724 (1988).
17. Ma, J. & Ptashne, M. *Cell* **51**, 113-119 (1987).
18. Giniger, E. & Ptashne, M. *Nature* **330**, 670-672 (1987).
19. Lech, K., Anderson, K. & Brent, R. *Cell* **52**, 179-184 (1988).
20. Celenza, J. L. & Carlson, M. *Science* **233**, 1175-1180 (1986).
21. Hoeffler, W. K., Kovelman, R. & Roeder, R. *Cell* **53**, 907-920 (1988).
22. Sorger, P. K. & Pelham, H. R. B. *Cell* (in the press).
23. Godowski, P. J., Picard, D. & Yamamoto, K. R. *Science* **241**, 812-816 (1988).
24. Webster, N., Green, S., Jin, J. J. & Chambon, P. *Cell* **54**, 199-207 (1988).
25. Hollenberg, S. M. & Evans, R. *Cell* (submitted).
26. Schena, M. & Yamamoto, K. *Science* **241**, 965-968 (1988).
27. Metzger, D., White, J. H. & Chambon, P. *Nature* **334**, 31-36 (1988).
28. Fischer, J. A., Giniger, E., Maniatis, T. & Ptashne, M. *Nature* **332**, 853-856 (1988).
29. Ma, J., Przibilla, E., Hu, J., Bogorad, L. & Ptashne, M. *Nature* **334**, 631-633 (1988).
30. Kakidani, H. & Ptashne, M. *Cell* **52**, 161-167 (1988).
31. Webster, N., Jin, J. R., Green, S., Hollis, M. & Chambon, P. *Cell* **52**, 169-178 (1988).
32. Serfling, E., Jasin, M. & Schaffner, W. *Trends Genet.* **1**, 224 (1985).
33. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. *Nature* (in the press).
34. Marsden, J. C. *J. Virol.* **61**, 2428-2437 (1987).
35. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. *Genes Dev.* **2**, 718-729 (1988).
36. Triezenberg, S. J., LaMarco, K. L. & McKnight, S. L. *Genes Dev.* **2**, 730-742 (1988).
37. O'Hare, P. & Goding, C. R. *Cell* **52**, 435-445 (1988).
38. Preston, C. M., Frame, M. C. & Campbell, M. E. *Cell* **52**, 425-434 (1988).
39. McKnight, J. L., Kristie, T. M. & Roizman, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7061-7065 (1987).
40. Maniatis, T., Goodbourn, S. & Fischer, J. A. *Science* **236**, 1237-1245 (1987).
41. Berk, A. J. *A. Rev. Genet.* **20**, (eds Campbell, A., Herskowitz, I. & Sandberg, L. M.) 45-81 (1986).
42. Random House Dictionary of the English Language (Random House, New York, 1967).
43. Struhl, K. *Molec. cell. Biol.* **6**, 3847-3853 (1986).
44. Schaffner, A., Schirm, S., Muller-Baden, B., Welsch, F. & Schaffner, W. *J. molec. Biol.* **201**, 81-90 (1988).
45. Ondek, B., Gloss, L. & Herr, W. *Nature* **333**, 40-45 (1988).
46. Herr, W. & Clarke, J. *Cell* **45**, 461-470 (1986).
47. Davidson, I., Xiao, J. H., Rosales, R., Staub, A. & Chambon, P. *Cell* (in the press).
48. Fromental, C., Kanno, M., Nomiya, H. & Chambon, P. *Cell* (in the press).
49. Giniger, E. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 382-386 (1988).
50. Lin, Y. S., Carey, M. F., Ptashne, M. & Green, M. R. *Cell* **54**, 659-664 (1988).
51. Lee, K. A. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8355-8359 (1987).
52. Sawadogo, M. & Roeder, R. G. *Cell* **43**, 165-175 (1985).
53. Carthew, R. W., Chodosh, L. A. & Sharp, P. A. *Cell* **43**, 439-448 (1985).
54. Sigler, P. B. *Nature* **333**, 210-212 (1988).
55. Horikoshi, M., Hai, T., Lin, Y. S., Green, M. R. & Roeder, R. G. *Cell* (in the press).
56. Horikoshi, M., Carey, M. F., Kakidani, H., Roeder, R. G. *Cell* **54**, 665-669 (1988).
57. Buratowski, S., Hahn, S., Sharp, P. A. & Guarente, L. *Nature* **334**, 37-42 (1988).
58. Cavallini, B. *et al. Nature* **334**, 77-80 (1988).
59. Fischer, J. A. & Maniatis, T. *Cell* **53**, 451-461 (1988).
60. Denis, M., Poellinger, L., Wikstrom, A. C. & Gustafsson, J. A. *Nature* **333**, 686-688 (1988).
61. Pratt, W. B. *et al. J. biol. Chem.* **263**, 267-273 (1988).
62. Picard, D. & Yamamoto, K. *EMBO J.* **6**, 3333 (1987).
63. Howard, K. J. & Distelhorst, C. W. *J. biol. Chem.* **263**, 3474-3481 (1988).
64. Ma, J. & Ptashne, M. *Cell* **50**, 137-142 (1987).
65. Johnston, S. A., Salmeron, J. M. Jr & Dincher, S. S. *Cell* **50**, 143-146 (1987).
66. Lue, N. F., Chasmin, D. I., Buckman, A. R. & Kornberg, R. D. *Molec. Cell. Biol.* **7**, 3446-3451 (1987).
67. Ma, J. & Ptashne, M. *Cell* (in the press).
68. Baeuerle, P. A. & Baltimore, D. *Cell* **53**, 211-217 (1988).
69. Magasanik, B. *TIBS* (in the press).
70. Minchin, S. D., Austin, S. & Dixon, R. A. *Molec. Microbiol.* **2**, 433-442 (1988).
71. Ninfa, A. J., Reitzer, L. J. & Magasanik, B. *Cell* **50**, 1039-1046 (1987).
72. Keener, J. & Kustu, S. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4976-4980 (1988).
73. Helmann, J. D. & Chamberlin, M. J. A. *Rev. Biochem.* (in the press).
74. Hochschild, A. & Ptashne, M. *Cell* **44**, 925-933 (1986).
75. Griffith, J., Hochschild, A. & Ptashne, M. *Nature* **322**, 750-752 (1986).
76. Hochschild, A. & Ptashne, M. *Nature* (submitted).
77. Giniger, E., Varnum, S. M. & Ptashne, M. *Cell* **40**, 767-774 (1985).
78. Bram, R. J. & Kornberg, R. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 43-47 (1985).
79. Jencks, W. P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4046-4050 (1981).
80. Sabbah, M., Reduilh, C., Secco, C. & Baulieu, E. E. *J. biol. Chem.* **262**, 8631-8635 (1987).
81. Kumar, V. & Chambon, P. *Cell* (in the press).
82. Goute, C. & Johnson, A. D. *Cell* **52**, 879-882 (1988).
83. Hollis, M., Valenzuela, D., Pioli, D., Wharton, R. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5834-5838 (1988).

Intermediate-term prediction of occurrence times of strong earthquakes

V. I. Keilis-Borok*, L. Knopoff†, I. M. Rotwain* & C. R. Allen‡

* Institute of the Physics of the Earth, Academy of Sciences of the USSR, Moscow, USSR

† Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

‡ Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA

Pattern recognition procedures for infrequent events are adapted to the problem of identifying patterns of clustering of small- and intermediate-scale seismicity before large earthquakes. Identification procedures derived from analysis of large California and Nevada earthquakes yield a high success rate when applied to other parts of the world.

The problem of earthquake prediction can be posed as a successive contraction of the space-time domain within which an earthquake is expected to occur. The identification of large-scale regions where earthquakes are likely to occur, without regard to the times of their occurrence, may be considered to be the initial stage of the succession¹⁻⁵. The next stage involves the identification of relatively large areas where strong earthquakes can be expected to occur within time intervals of the order of decades; the seismic gap model is an example of current progress in this phase^{6,7}. Although there is some progress on the prediction of small- and intermediate-sized earthquakes at short time- and distance-scales^{8,9}, there is very little progress on the prediction of large earthquakes. In this paper we explore the possibility of identifying the times of increased probability (TIP) of occurrence of a strong earthquake on an intermediate timescale of the order of a few years and on a distance scale of the order of several hundred kilometres. We are not aware of any efforts to unify predictions over the three timescales, nor do we attempt any unification here. The prediction of the Haicheng (1975) earthquake may represent an encouraging first-effort case history at such unification¹⁰.

Earthquake precursors

Numerous occurrences of natural phenomena preceding strong earthquakes have been reported^{11,12}. The phenomenology spans such diverse fields as geophysics, geochemistry and animal behavior. But, only a small number of the proposed precursors are sufficiently well defined to be tested statistically. As far as we know, the statistical significance of only one intermediate-term precursor has been evaluated; this precursor was found to be statistically acceptable¹³. The problem as a whole is far from being solved either theoretically or practically. Even in as large an active region as California, we have neither a sufficient occurrence of strong earthquakes to which customary statistical methods can be applied, nor theories for their occurrence (which could be used to reduce the size of the database necessary for statistical analysis). Therefore we have recourse to an alternative approach to intermediate-term prediction, using methods of pattern recognition in infrequent events. These procedures have been described in detail by Gelfand *et al.*¹. Pattern recognition methods have been applied to the space prediction of strong earthquakes¹⁻⁵; here we apply them to time prediction.

We restrict our attention to predictions based on the space-time variation of seismicity, as this database is probably the most complete, in comparison with any other, over long periods of time for many regions of the world. Even with these data, identifications of a single type of anomalous seismicity may be unreliable because of the wide variety of local environments in which earthquakes occur and/or the fact that the dynamics of seismicity are nonlinear and hence potentially unstable, having an intrinsically chaotic or pseudo-chaotic component. Accord-

ingly, premonitory patterns of seismicity may be rather different from one strong earthquake to another.

To address this difficulty we base our predictive model on a broadly chosen set of features, without imposing the condition *a priori* that any single feature constitutes a precursor that is sufficient for prediction. We call the quantitative definitions of our candidate precursory features *traits*. A trait is a property of the sequence of earthquakes of a region in a finite sliding time-window and a certain magnitude range. In this way a point sequence of earthquakes in the time-space-magnitude domain is smoothed into a less erratic representation of temporal seismicity. Each trait has been given a broad but unambiguous definition.

We try to identify TIPs before strong earthquakes using appropriate combinations of 12 traits of earthquake sequences. The rule for identification is determined by methods of pattern recognition of infrequent events. Strong earthquakes are defined as main shocks with magnitudes greater than or equal to some threshold M_0 . Main shocks are themselves defined as the residue of a catalogue after aftershocks have been removed to avoid giving too much weight to the individual events in a single cluster of earthquakes.

The 12 traits that we have considered depict activity (A), quiescence (Q), temporal rates of change of seismic activity (T), sizes of earthquakes (S), spatial and temporal clustering (C) and long-range interactions between earthquakes (I). A qualitative description of these traits is the following:

(A) Activity is measured by the number of earthquakes in given time and magnitude intervals, weighted or not according to their energies.

Table 1 Parameters in pattern recognition

Function	n^*	s (years)	Other parameters	S-M-L thresholds
A1b	0.4	3	0†	
A1c	1.4	3	$y = 7$ yr	3, 5
A2	3.0	3	$c = 0.1$	36, 71
A3	3.0	1	$r = 0.5$	0.5, 0.67
Q1	1.4	6	$p = 50\%$; $y = 3$ yr	0, 12
T2	1.4	2		-1, 1
S1	3.0	1	$c = 0.1$; $y = 2$ yr	4.1, 4.6
S2	3.0	1	$c = 0.1$; $y = 3$ yr	7.9, 14.2
C1		3	$e = 2$ days (ref. 15) $M_a = 2.8$; $M_{c0} = 4.5$	12, 24

* The cutoff magnitudes for a given n are approximately:

n	3	1.4	0.4
M	4.9	5.3	6.0 (north) 6.2 (south)

† S-L only.

(Q) Quiescence identifies low activity when compared with the average.

(T) Temporal rates of change are measured as deviations from long-time trends or as a difference of activity between two consecutive time intervals.

(S) Size of earthquakes is either the average area of an earthquake source or the average ratio of the linear dimension of the source to the distances between sources.

(C) Clustering is measured by the number of aftershocks of main shocks in the medium-magnitude range.

(I) Long-range interaction is a measure of the simultaneous occurrence of main shocks at relatively large distances.

Formal definitions are given in the next section.

This list spans most if not all of the premonitory seismicity patterns that have been proposed on either empirical or theoretical grounds. The traits are normalized by the choice of a lower earthquake magnitude cutoff M_c so that our procedures can be applied uniformly to any seismic region. For eight of the traits, M_c is defined by the condition that the average annual number of main shocks be equal to a specified value n ; values of the lower-magnitude cutoff that correspond to various values of n for California are given in Table 1. Four traits are normalized by assuming that M_c is a fixed-magnitude interval c below M_0 . Other less general conditions are also given below. All traits except one are therefore smoothed functions of the sequence of main shocks.

Trait definitions

The definitions of the traits we use are as follows:

Seismic activity (A). Let E_i be the energy released in the i th earthquake, and β be an exponent between 0 and 1. Our measure of seismic activity $E_\beta(t, s; n)$, is the sum of the quantities E_i^β in a sliding window extending from time $(t-s)$ to t . The lower cutoff magnitude is determined from the value of n , which is the average number of main shocks per year; for values of cutoff magnitudes see Table 1. If $\beta=0$, we count the number of earthquakes in the window; if $\beta=2/3$, we have a measure of the total area of fault breaks¹⁴ (see applications in refs 15 and 16. If $\beta=\frac{1}{3}$, we have a measure of the sum of the linear dimensions of all fault breaks; if $\beta=\frac{1}{2}$, we obtain the windowed Benioff strain. We used the relation $\log_{10} E_i = 1.5 (M_i - 4.5)$, in which the coefficient 4.5 is arbitrary and is chosen for convenience. Four useful versions of the activity trait are: the number of main shocks $(A1) = E_0$; the number of main shocks y years earlier $(A1c) = E_0(t-y)$; the area of fault breaks $(A2) = E_{2/3}$, $M(n) < M < M_0 - c$; the relative number of main shocks in a finite magnitude interval $(A3) = 1 - [E_0(t, s; rn)/E_0(t, s; n)]$, where $r < 1$. We have used $r = \frac{1}{2}$ (Table 1), in which case the magnitude interval is close to 0.3 units.

Quiescence (Q). Let $E_{0,p}$ be the lower p th quantile of the function E_0 . Then, a first measure of quiescence (Q1) is the amount of activity below the p th quantile, and is given by

$$Q1 = \int_{t-s-y}^{t-y} \text{Pos}[E_{0,p} - E_0(t, s; n)] dt$$

where $\text{Pos}[x]$ denotes positive values of x only. The function Q1 may serve to identify the existence of a seismic gap.

A second measure of quiescence, Q2, can be defined as follows. We let t_x be the time of the most recent maximum of the function E_0 , and t_n be the time of a minimum of E_0 between t_x and t . Then, Q2 is the sum of the two positive quantities,

$$Q2 = [E_0(t_x) - E_0(t_n)] + [E_0(t) - E_0(t_n)]$$

If there is no intervening minimum in $E_0(t)$ between t_x and t , we set $t_n = t$.

Rate of temporal variation of activity (T). Consider the cumulative number of main shocks in the time interval since some fixed time t_0 ; this function is $E_0(t, s = t - t_0)$. As a measure of the deviations from long-term trends, T1, we use the difference

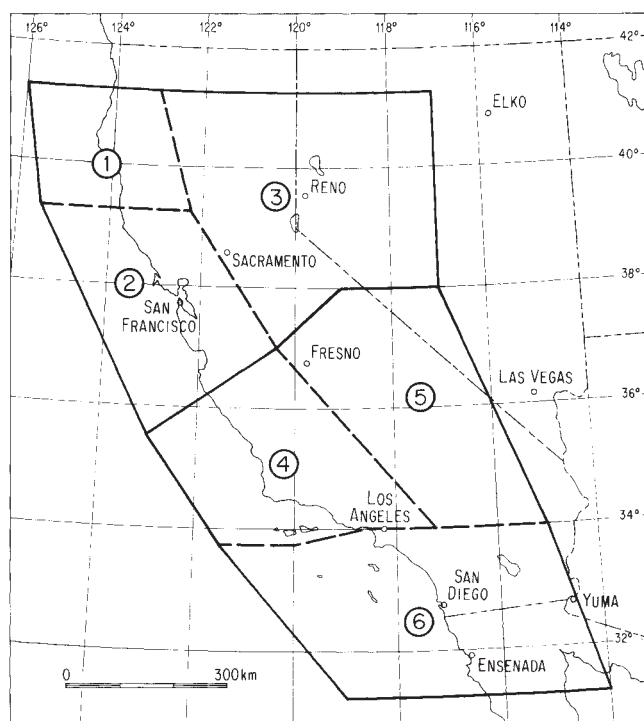


Fig. 1 Regions (solid boundaries) and areas (dashed boundaries) of California and Nevada.

between these cumulative quantities at times t and $(t-y)$, $y > 0$:

$$T1 = E_0(t, t - t_0) - E_0(t-y, t - t_0 - y) \frac{(t - t_0)}{(t - y - t_0)}$$

The second term is normalized to the same time interval as the first.

We also consider the difference between the values of E_0 at times separated by the window length; that is, we define an increase of activity over the interval s to be $(T2) = E_0(t) - E_0(t-s)$.

Size of earthquake sources (S). Macrofailure in rocks is assumed to occur when the ratio between the average linear dimension of microsources to the average linear distance between them exceeds some threshold¹⁷. We consider the maximum of an estimate of this ratio within the last y years. A measure of the proximity to macrofailure (S1) is given by

$$S1 = \text{Max}_s \left[\frac{1}{E_0^{2/3}} E_{1/3} \right], M(n) < M < M_0 - c$$

where $\text{Max}_s[x]$ is the largest value of x in the s -window.

According to some models¹⁸, the magnitudes of main shocks should increase before a strong earthquake. As a diagnostic of this effect, we define the average area of rupture surfaces of sources (S2) as

$$S2 = \text{Max}_y \left[\frac{1}{E_0} E_{2/3} \right]$$

Spatial and temporal clustering (C). The occurrence of abnormally large clusters of aftershocks of main shocks has been proposed as a precursor of strong shocks^{15,16} and is the only precursor to date whose statistical significance has been validated¹³. We define B_i as the number of aftershocks with magnitudes above some threshold M_a that occur after the i th main shock, during the interval from t_i to $t_i + e$. Then, the diagnostic C1 is $\text{Max}_i[B_i]$, that is, the largest value among the quantities B_i for main shocks in the window s .

Long-range interactions (I). The neighbourhood of a forthcoming strong earthquake may have an increase in activity after another strong earthquake has occurred at a relatively large

distance¹⁹. As a measure of this increase, we consider the magnitudes of main shocks that occur within a time interval y after a strong earthquake has occurred in the same or an adjacent region. These events are called long-range aftershocks. The diagnostic I1 is $\text{Max}_y[M_I]$ where M_I is the maximum magnitude of the long-range aftershocks, which are themselves main shocks.

A second measure of long-range interaction involves the contemporaneous change of activity in adjacent regions, or if we consider a trait of an area, which is a smaller unit of territory than a region, the activity in the region to which the area belongs. The measure I2 is defined as $E_0(M_0 - c) - E_0(M_0)$ in the region (or super-region).

Data sources

We divide California and western Nevada into northern and southern regions along a boundary in the neighbourhood of 37° N (Fig. 1). We assign TIPs to one or the other of these regions separately. To see if the TIPs can be narrowed to smaller spatial units, we also subdivide each of the regions into areas. In our opinion, the regions and areas defined in Fig. 1 have objective and unifying seismotectonic characteristics²⁰⁻²⁴; they were defined before analysis. It is tempting to assign the TIPs to as small a territory as possible, but the regions and areas must not be too small, for two reasons. First, strong-earthquake precursors may be spread out over a large territory, especially at the stage considered here, so that they will be difficult to identify locally; second, for small territories, the number of earthquakes would be insufficient to carry out our procedures. We recognize that a prediction covering a smaller territory would be scientifically more meaningful and socially more useful than that which we are able to provide.

We identify TIPs in the southern region from 1938 onward from analysis of the CIT²⁵ catalogue and in the northern region from the NOAA catalogue²⁶. We removed aftershocks from the catalogues by the procedure given in ref. 15. The remaining or main shocks from about 50% of the southern catalogue and about 85% of the northern catalogue at these cutoff magnitudes.

We choose strong earthquakes to have $M > M_0 = 6.4$ (ref. 15). Ten strong earthquakes occurred in the southern region and six in the northern region since 1938. Of these, the two Long Valley shocks of 1980 in the southern region occur within one day of each other; we consider these two earthquakes as a single event. Because only five days elapsed between the strong Fallon, Nevada and Cape Mendocino events of December 1954, both in the northern region, we cannot identify which of the two would be anticipated by a TIP for the entire northern region that begins some months or years before this date; since the two events occur in different areas we may try to assign a TIP to each of these areas.

Data analysis

We consider the vector $\mathbf{P}_m(t) = \{f_a(t)\}$, where f_a is one of the sequence of traits listed above, with parameters as in Table 1, and m is the index of the region or area. Given $\mathbf{P}_m(t)$, our task is to decide whether a TIP is manifested in the subsequent time interval $(t, t + \tau)$ in the m th territory.

We used the 'subclasses' pattern-recognition algorithm¹. The first, or 'learning', stage of this procedure gives the rule for identification of TIPs. For this purpose, the history of each territory is divided into two parts: the D (or 'dangerous') time intervals, which immediately precede each strong earthquake and therefore include TIPs, and the N (or 'non-dangerous') time intervals, which occupy the remainder of time. The duration of the D intervals was taken to be two years. The vector $\mathbf{P}_m(t)$ is defined both inside and outside the TIPs. The three-year intervals of N-type after each strong earthquake were not used in learning, to avoid complications connected with high residual activity, which is not completely suppressed by removal of aftershocks.

In an effort to make the range of values of each function

Table 2 Features of time intervals

a Time intervals D $K_1 = 7; \bar{K}_1 = 2^*$									
No.	A1b	A1c	A2	A3	Q1	T2	S1	S2	C1
1						M	L		
2							L		
3a						L		L	
3b						L		S	
4						S			L
5a						L	L		S
5b						S	L		S
6					S		L		
7					L				
8		S			S				S
9		S				S			S
10				S		S			S
11			S			L			
12				L		L			
13			L			L	L		
14	L					L	S		
b Time intervals N $K_2 = 10; \bar{K}_2 = 4^*$									
No.	A1b	A1c	A2	A3	Q1	T2	S1	S2	C1
1		S	L		L				
2			L		L				L
3					L	L			L
4		S							L
5a								S	L
5b								L	S
6	S						L		L
7						L	L		L
8							L		L
9		L					L		
10						S			
11			S			L			
12a						L		S	
12b						S		L	
13					L			S	
14		S						S	
15				L		L		L	
16			L			S			
c Frequency of appearance among features									
A1b	A1c	A2	A3	Q1	T2	S1	S2	C1	
2	6	6	3	8	20	12	10	14	

* Defined as in ref. 1 for the algorithm 'subclasses'. Each D feature is present in $\geq K_1$ groups of vectors N (subclasses) and $\leq$ in K_1 vectors D. K_2 and $\bar{K}_2$ are similarly defined for N features.

robust, we discretize it by dividing it into three intervals, small (S), medium (M) and large (L); one function, A1b, is divided into S and L intervals only. Each interval contains an approximately equal number of total (D + N) sample times; thus the D/N ratio in each interval depends as little as possible on *a priori* knowledge of the times of strong earthquakes. The value of each function is represented by a binary code that indicates whether it is in the S, M or L range. In this way the complete vector $\mathbf{P}_m(t)$ is represented by a short binary code which is the set of yes/no answers about the discretized values of the functions. The sampling is made at the 14 D-times immediately before each strong earthquake and 24 N-times distributed more or less evenly among the N intervals. The distributions of the values of each function are compared for sample times D and N separately. If the difference between the distributions is sufficiently large, the function is used for further analysis. The functions listed in the section on trait definitions have passed this screening. The fixed parameters and thresholds used for discretization are listed in Table 1; these choices are justified elsewhere²⁷.

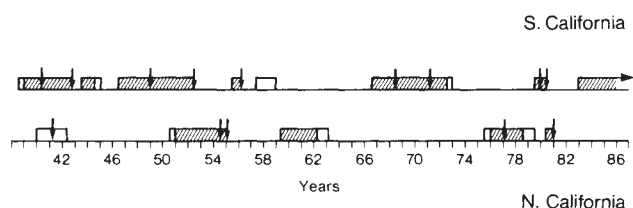


Fig. 2 TIPs and strong earthquakes in California-Nevada. The earthquake of 9 February 1941 is not preceded by a TIP if $\bar{\Delta} = 5$. There are three false alarms. Arrows correspond to the times of strong earthquakes ($M > 6.4$). Shaded segments denote TIPs with $\bar{\Delta} = 5$; unshaded segments denote TIPs with $\bar{\Delta} = 4$. A TIP began in 1983, but because the catalogue available to us terminated in April 1985, the 1983 TIP could only be extended until April 1986. We do not know as yet if this TIP extended an additional three months and hence preceded the Chalfant Valley earthquake of 21 July 1986, but the association is a plausible one.

The vectors $\mathbf{P}_m(t)$ for D and N intervals are compared to define the D and N features. Each feature specifies the range for one of the traits, or for two or three of the traits together. Because we do not know the lead time for the emergence of premonitory patterns, we considered three samples of each D interval, namely just before an earthquake and one and two years before; these samples constitute a subclass¹. A D feature is selected under the condition that it appears sufficiently often in D intervals and sufficiently rarely in N intervals, and *vice versa* for N features. A complete description of the algorithm is given in ref. 1. The analysis for regions makes use of eight traits from the full list. The D and N features for the two regions are shown in Table 2; the notation $\bar{S}$ and $\bar{L}$ signifies that the trait is characterized as 'not S' or 'not L'.

In general, a trait that appears frequently in a list of features influences recognition strongly. The frequency of appearance of traits is listed in Table 2c. Contrary to our expectations, activity as measured by the number of main shocks is not a major diagnostic, whereas large changes of seismicity, large numbers of aftershocks and an increase in the dimensions of the sources are indeed important. Generally, the functions T2, C1, S1, S2, Q1 and A2 are not large in N features, while A1c is not small, and the opposite statements apply to D features.

There are four exceptions to internal consistency out of the 76 entries in Tables 2a and b (excluding traits A1b and A3). If a trait appears as L or $\bar{S}$ in D intervals, it is most likely to appear as $\bar{L}$ or S in the N regions. Two of the features in the N intervals and two in the D intervals have the same combinations of traits.

Qualitatively these results suggest that strong earthquakes are typically preceded within one to three years by an increase in activity (T2), high clustering (C1) and high activity as represented by the concentration (S1) and average size of main shocks (S2), although not all of these need occur in all cases. Efforts to find simpler diagnostics than those in Table 2 are explored by Gabrielov *et al.*²⁸.

Diagnosis of TIPs

To identify D times and thus to diagnose TIPs, we count the number of D and N features at sample times over consecutive two-month intervals in each region; for practical purposes this is a continuous sampling. Let $n_D(t)$ and $n_N(t)$ be the numbers of D and N features at time t and define $\Delta(t) = n_D - n_N$. If Δ is sufficiently large, the particular time is likely to be a D interval. The rule for the identification of D times is that: [a TIP starts when $\Delta \geq \bar{\Delta} = 5$ and $A2 \leq \bar{S} = 124$; a TIP lasts for $\tau = 12$ months]. In the evaluation of A2 for the latter condition, no limitations are imposed on the maximum magnitude of main shocks. This condition on A2 implies that a strong earthquake is not likely to occur if sufficiently large parts of the fault system have been recently unlocked, whether by a strong earthquake or by several

Table 3 Performance of algorithm for identification of TIPs

	$\bar{\Delta} = 4$		$\bar{\Delta} = 5$	
	Duration	Failures to predict/ successful TIPs	Duration	Failures to predict/ successful TIPs
Northern California	32%	0/5	22%	1/5
Southern California	52%	0/9	46%	0/9
Caucasus	15%	1/4	9%	2/4
Central Asia	39%	1/9	27%	2/9
Northern Appalachians			25%	0/2
Brabant-Ardenne			25%	0/3
Vrancea			25%	0/2

earthquakes of smaller magnitudes; the quantity $\bar{S} = 124$ is the equivalent of one earthquake with $M_L = 7.2$. The TIPs this diagnosed (Fig. 2) occupy about one-third of space-time and precede 13 out of 14 strong earthquakes. The sole failure to predict is a 1941 earthquake in northern California, and even this is predicted if we reduce $\bar{\Delta}$ to 4 (Table 3). Evidently the small number of failures to predict is predetermined by the learning procedure and the retrospective choice of the thresholds $\bar{\Delta}$, $\bar{S}$ and τ . These results must be therefore tested on independent data that were not involved in the learning phase.

Stability of the results was tested by elimination of each of the nine diagnostic functions in turn from the learning stage of our algorithm. Only the elimination of T2 increases the number of false alarms significantly; all of the others can be eliminated singly without major damage to the results of the analysis. There are no major changes in the results if: (1) we eliminate each of the strong earthquakes from learning, one at a time, or if (2) we use only the first chronological 60% of the earthquake catalogue (see similar test in ref. 1). These stability tests are described in ref. 28.

Application to other regions

Thirteen of the 14 strong earthquakes in California and Nevada between 1938 and 1983 were preceded by TIPs. The TIPs arise because of the appearance of certain combinations of patterns of seismicity before these earthquakes. It is not clear whether a procedure derived retrospectively and with a large number of adjustable parameters is applicable to the prediction of future strong earthquakes. To answer this question, one must test whether our list of features for the identification of TIPs in California and Nevada can be applied without significant modification to other seismic regions of the world.

A direct transfer of the algorithm to an independent set of earthquake data was made specifically for the Pamir-Tien Shan and the Caucasus seismic zones of the USSR from 1962 to 1983. These two zones are somewhat similar to California and Nevada, at least in their neotectonic setting and seismic activity. Nine of the 13 strong earthquakes that occurred in these regions from 1966 to 1984 were preceded by TIPs at the threshold $\bar{\Delta} = 5$, and 11 out of 13 were preceded by TIPs at the threshold $\bar{\Delta} = 4$. The durations of these TIPs are indicated in Table 3.

We also transferred the algorithm without change to several regions with very different tectonic setting and levels of seismicity from the above two regions. In the intraplate regions of the Northern Appalachians (1963–1985) the strong earthquakes are defined by $M_0 = 5$ and in the Brabant-Ardenne massif of Belgium (1963–1985), $M_0 = 4.5$. In the Vrancea region of the Eastern Carpathians (1966–1985), the strong earthquakes have focal depths down to 100 km and magnitudes $M_0 = 6.4$. During the periods for which we had data, seven strong earthquakes with $M > M_0$ occurred in all three regions and all of these were

preceded by TIPs with a duration of 25% of the total interval. There were three false alarms.

A similar test was performed on an even wider suite of seismic zones with another version of the algorithm that differs from the above only in the values of certain parameters²⁸. Twenty-three out of 29 strong earthquakes that occurred in these zones were preceded by TIPs occupying 25% of the total elapsed time.

For brevity, we bypass a detailed description²⁸ of the assignment of a regional TIP to one of the smaller areas within a region. The most frequently occurring trait is Q2, which is large in D features and small in N features. The number of exceptions to regularity of traits is significantly larger among the features in this case than among the diagnostics for regional prediction. Except for trait T2, there are no significant dissimilarities between the areas and the regions. The total space-time occupied by the 15 TIPs in the areas is now only 13%, but at the cost of

two more failures-to-predict (the Imperial Valley earthquakes of 1942 and 1956). In some cases TIPs are diagnosed in two or three areas simultaneously instead of only one. Conclusions for areas are speculative at this time since we have not subjected the algorithm for areas to the tests of stability and of transfer to other parts of the world.

The statistical significance of the observations cannot be estimated because of the small sample size. Furthermore, we have not formalized methods for the choice of regionalization nor for the choices of many of the fixed parameters. We do not claim that the algorithm is optimal. But the success of the geographical transfers for the regional studies suggests that the set of traits we have identified is intrinsically relevant to the approach of a strong earthquake in diverse tectonic settings and that the algorithm deserves the decisive test of the prediction of future strong earthquakes.

Received 12 April; accepted 16 August 1988.

1. Gelfand, I. M. *et al. Phys. Earth planet. Inter.* **11**, 227-283 (1976).
2. Briggs, P., Press, F. & Guberman, Sh. A. *Bull. geol. Soc. Am.* **88**, 161-173 (1977).
3. Caputo, M. *et al. Phys. Earth planet. Inter.* **21**, 305-320 (1980).
4. Gvishiani, A. D. & Soloviev, A. A. *Earthquake Prediction Res.* **2**, 237-243 (1984).
5. Jones-Cecil, M., Wheeler, R. L. & Dewey, J. D. *U.S. geol. Sur. Open File Rep.* 81-195 (1981).
6. McCann, W. R., Nishenko, S. P., Sykes, L. R. & Krause, J. *PAGEOPH* **117**, 1082-1147 (1979).
7. Sykes, L. R. & Nishenko, S. P. *J. geophys. Res.* **89**, 5905-5927 (1984).
8. Jones, L. M. *Bull. seismol. Soc. Am.* **75**, 1669-1679 (1985).
9. Kagan, Y. Y. & Knopoff, L. *Science* **236**, 1563-1567 (1987).
10. Raleigh, C. B. *EOS* **58**, 236-272 (1977).
11. Mogi, K. *Earthquake Prediction* (Academic, Tokyo, 1985).
12. Rikitake, T. *Earthquake Prediction* (Elsevier, Amsterdam, 1976).
13. Molchan, G. M., Dmitrieva, O. E. & Rotvain, I. M. *Comput. Seismol.* **21** (in the press).
14. Keilis-Borok, V. I. & Malinovsky, L. N. *J. geophys. Res.* **69**, 3019-3024 (1964).
15. Keilis-Borok, V. I., Knopoff, L., Rotvain, I. M. & Sidorenko, T. M. *J. geophys. Res.* **85**, 803-811 (1980).
16. Keilis-Borok, V. I., Knopoff, L. & Rotvain, I. M. *Nature* **283**, 259-263 (1980).
17. Sobolev, G. A. & Zavalov, A. D. *Dokl. Akad. Nauk SSSR* **252**, 69-71 (1980).
18. Kagan, Y. Y. & Knopoff, L. *J. geophys. Res.* **86**, 2853-2861 (1981).
19. Prozorov, A. G. *Comput. Seismol.* **14**, 20-26 (1983).
20. Jennings, C. W. *Calif. Div. Mines Geol. Geologic Data Map No. 1* (1975).
21. Jennings, C. W. *Calif. Div. Mines Geol. Geologic Map of California* (1977).
22. Real, C. R., Topozada, T. R. & Parke, D. L. *Calif. Div. Mines Geol. Map Sheet 39* (1978).
23. Slemmons, D. B., Gimlett, J. I., Jones, A. E., Greensfelder, R. & Koenig, J. *Nevada Bur. Mines Map 29* (1965).
24. Stover, C. W. *U.S. Geol. Sur. Geophys. Inv. Map 984* (1986).
25. *Earthquakes in Southern California, 1932-1985* (Seismol. Lab., Calif. Inst. Tech., 1976-1986).
26. *World hypocenter data file 1932-1985* (US geol. Surv., natn Earthquake Inf. Serv., 1986).
27. Keilis-Borok, V. I., Lamoreaux, R., Johnson, C. & Minster, B. *Earthquake Prediction Res.* **1**, 135-151 (1982).
28. Gabriellov *et al.*, Preprint, Centre Regional de Sismologia para America del Sur, Lima.

NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A

Jayant B. Udgaonkar & Robert L. Baldwin

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA

The presence of an early intermediate on the folding pathway of ribonuclease A has been demonstrated by a study of the exchange reaction between the backbone amide protons in the folding protein and solvent protons using rapid mixing techniques. A structural analysis of the intermediate by two-dimensional ¹H-NMR is consistent with the framework model of protein folding in which stable secondary structure first forms the framework necessary for the subsequent formation of the complete tertiary structure.

To obtain detailed structural information about the folding pathway of a protein has been a challenge of long standing. The single-domain protein ribonuclease A (RNase A) has traditionally served as an archetypal model for small proteins in protein folding and its folding pathway has been extensively studied. Multiple unfolded forms, which differ in their refolding rates, have been identified¹. There is strong evidence for the presence of both an early hydrogen-bonded intermediate^{2,3} and a late native-like intermediate^{4,5} on the folding pathway of the main unfolded species, but little is known about the structure or stability of these kinetic intermediates. Such information should provide the most useful evidence for establishing the folding pathway of RNase A.

Here we report the use of the exchange reaction of backbone amide protons with solvent protons to obtain detailed structural information about early events during the folding of RNase A. This experimental approach is also being used with another

small protein, cytochrome *c* (ref. 6). In our experiments, exchange is initiated by rapid (millisecond) mixing of a solution of RNase A in D₂O with an H₂O buffer at different times after initiation of refolding of the unfolded protein. The accessibility of individual backbone amide protons to exchange with solvent protons at each timepoint is measured by two-dimensional ¹H-NMR after folding is complete; it is this measurement that allows a structural characterization of the folding process. Our results reveal that a stable structural intermediate is formed within the first 1.5 seconds of folding. The rapid-mixing technique makes it possible to monitor the kinetics of formation of this intermediate, in which a large number of backbone amide protons are protected against exchange. NMR techniques have allowed assignment of several of these amide protons to specific residues found in the β -sheet of native RNase A. Thus, the β -sheet may be formed in the first major folding reaction of RNase A. The preliminary results reported here support the

framework model of protein folding⁷⁻⁹, in which the formation of stable secondary structure is a prerequisite to the formation of the final tertiary structure. The potential of the technique introduced here to provide more detailed structural information on protein-folding pathways is established.

Experimental strategy

Two-dimensional ¹H-NMR techniques¹⁰ allow the resolution of nearly all the backbone amide proton resonances in RNase A; for example, in a COSY (ref. 11) spectrum of RNase A in water, nearly all of the 120 or so expected C_αH-NH crosspeaks are observed in the fingerprint region (data not shown). The NMR techniques also make it possible to assign each such crosspeak to a specific backbone amide proton in the amino-acid sequence¹⁰. Twenty such assignments were reported earlier for RNase A (ref. 12), and additional assignments have been made in this laboratory (A. D. Robertson, unpublished results). The fact that the fingerprint region of the COSY spectrum of RNase A in H₂O is fully resolved implies that it will be possible to obtain a complete sequential assignment of all the crosspeaks in this part of the spectrum without too much difficulty. Recently, essentially complete sequential ¹H-NMR assignments have been reported for several proteins that are comparable in size to RNase A, including cytochrome *c* (ref. 13), lysozyme¹⁴ and thioredoxin¹⁵.

Elegant studies of the folding pathway of disulphide-reduced bovine pancreas trypsin inhibitor (refs 16 and 17) have shown that it is possible to trap covalently and then purify intermediates containing different disulphide bonds, and to study these intermediates using NMR techniques¹⁸. In the case of protein-folding reactions in which the disulphide bonds are left intact or otherwise are absent, it is not possible to trap folding intermediates covalently. Thus, although two-dimensional ¹H-NMR techniques can yield detailed structural information about a native protein the size of RNase A (ref. 10), there is a basic problem in obtaining similar information about the folding pathway of the disulphide-intact protein: the NMR techniques are slow (several hours are required to record a spectrum), whereas folding intermediates are typically populated only for short times (a few seconds or less). To use the capabilities of the NMR techniques, it is necessary first to label the structure of the folding intermediate in such a way that it can be studied in the native protein after folding is complete¹⁹⁻²². The exchange reaction of the backbone amide protons with solvent protons can be exploited for this purpose if the folding intermediate contains stable secondary structure (such as α -helices and β -sheets). The exchange rates of some of the backbone amide protons in a fully folded protein are slower than those in the unfolded protein by factors of as much as 10⁹, because these protons participate in hydrogen-bonded secondary structure in the fully folded protein^{23,24}. Hydrogen-bonded structure directly inhibits exchange because hydrogen bonds must be broken for exchange to occur²⁵. Accessibility to solvent is also required for exchange, and most highly protected amide protons in native proteins are both hydrogen-bonded and inaccessible to solvent. Thus at any time during the folding process, an amide proton may become resistant to exchange with solvent protons, either as a result of the involvement of that proton in secondary structure, or as a result of the seclusion of that proton from solvent, or both.

In the case of native RNase A, most of the 27 backbone amide protons that were protected against exchange with solvent deuterons over the one-year time period required for a neutron diffraction study^{26,27} are involved in hydrogen-bonding. These 27 amide protons are also distributed throughout the structure of the protein molecule^{26,27}. Under the conditions required to record the COSY NMR spectrum shown in Fig. 1, about 40 backbone amide protons are seen to be stable to exchange with solvent deuterons. Of these, 13 have been sequentially assigned

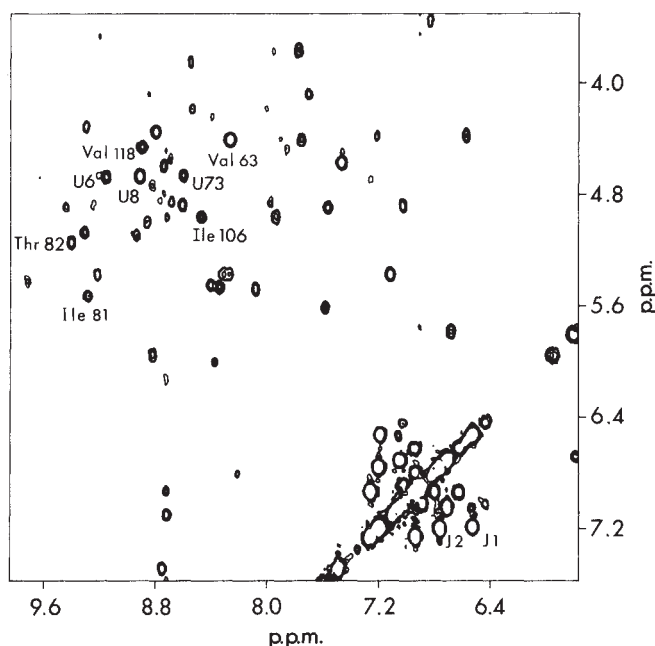


Fig. 1 Reference COSY spectrum of RNase A in D₂O, pH 4, 22 °C. Only the region showing C_αH-NH crosspeaks and aromatic ring-proton crosspeaks is depicted. The latter are shown at the lower right corner next to the diagonal of the spectrum. Only ~40 C_αH-NH crosspeaks are seen; the amide protons contributing to them are stable to exchange with solvent deuterons. Some of the sequential assignments for the C_αH-NH crosspeaks are shown. U6, U8 and U73 are C_αH-NH crosspeaks that have not yet been assigned to specific residues in the protein. The crosspeaks labelled J1 and J2 arise from completely non-exchangeable aromatic ring protons, and were chosen as internal intensity reference peaks (see Fig. 2 legend).

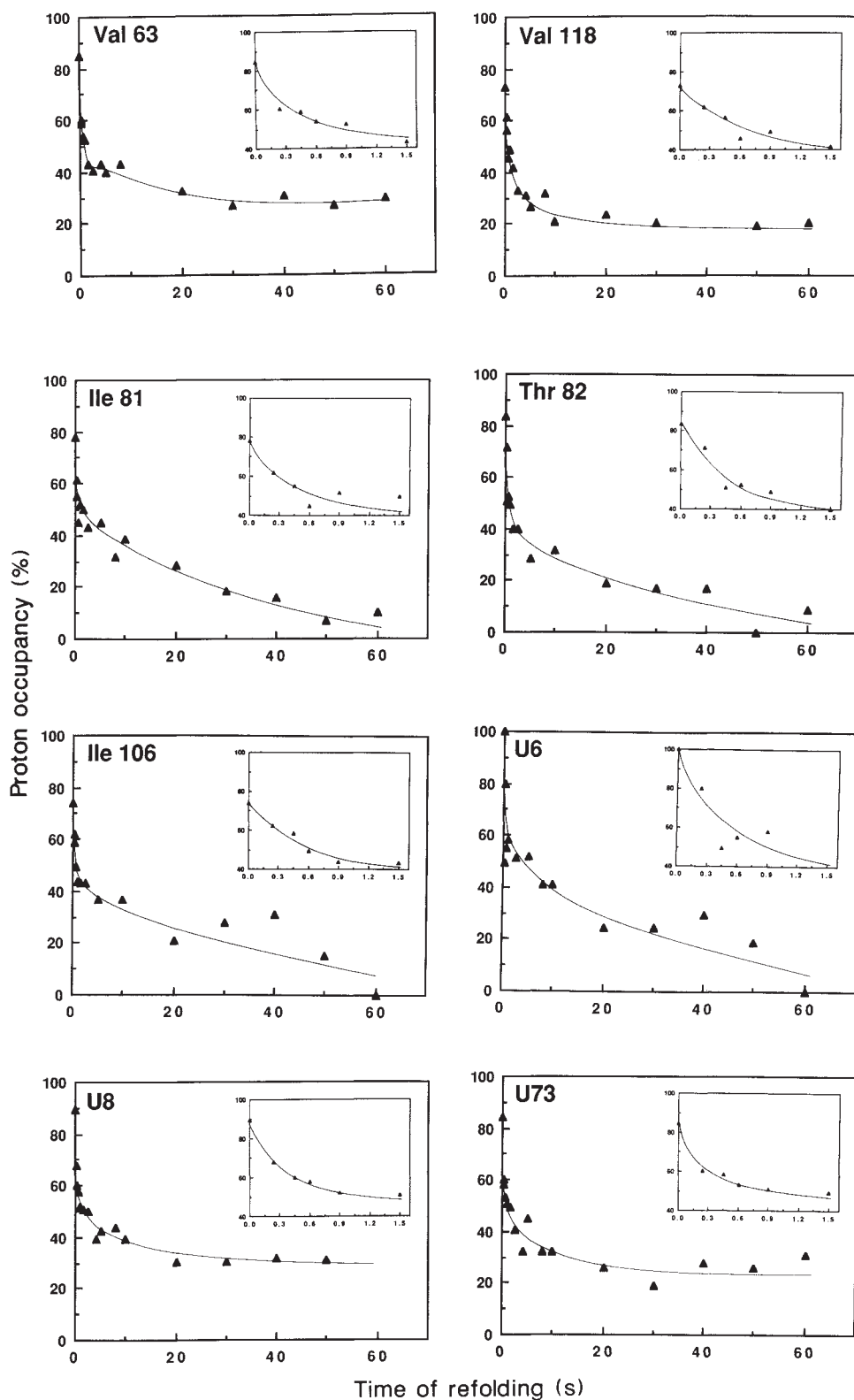
Methods. The preparation of the RNase A sample for the reference spectrum was as follows. RNase A (40 mg) were dissolved in 1.6 ml D₂O and 8.4 ml H₂O at pH 3 and the solution heated to 60 °C for 20 min; after lyophilization, the procedure was repeated to obtain protein that was 84% protonated at all amide hydrogen sites. The protein was then redissolved in 16% deuterated water at pH 3. The final NMR sample was prepared by ultrafiltration of the protein solution at pH 3, 5 °C; the solvent was completely deuterated by repeated concentration and then dilution of the sample in D₂O. The pH of the concentrated NMR sample (4 mM) in D₂O was raised to 4, and the sample was then kept at 5 °C for 20 h to allow exchange out of all protons labile to exchange. The two-dimensional homonuclear J-correlated (COSY) spectrum¹¹ was recorded at 500 MHz on a General Electric GN-500 spectrometer. The spectrum was obtained from 450 values of *t*₁ between 0 and 75.6 ms; each free induction decay consisted of 1,024 complex data points, and 64 scans were collected. Total data acquisition time was 10.5 h. The FTNMR program of Dennis Hare was used for data processing on a VAX 8550 computer. Before Fourier transformation, the time-domain data were multiplied by phase-shifted sinebell squared window functions, using phase shifts of 5° and 10° in the *t*₁ and *t*₂ dimensions respectively, and the *t*₁ domain data were extended to 1,024 points with zero-filling. An absolute magnitude calculation then yielded the final spectrum.

by NMR (ref. 12). All 13 are hydrogen-bonded in the crystal structure, according to the neutron diffraction data^{26,27}.

Two-dimensional ¹H-NMR techniques not only monitor the individual backbone amide protons, but also discriminate between protons and deuterons. If an amide proton site is initially deuterated when the protein is unfolded in D₂O, then the resistance (or accessibility) of the amide deuteron to exchange with solvent protons at a specific time during the folding process can be determined from the extent to which the site is protonated when H₂O is added at that time. If the amide proton is also stable to exchange in the native protein, the extent of proton-labelling can be determined after folding of the protein

Fig. 2 Kinetics of protection of backbone amide deuterons from exchange with solvent protons. The accessibility to exchange (proton occupancy) of 8 backbone amide proton sites is plotted versus time after initiation of refolding. The C_α -H-NH crosspeaks arising from these 8 amide protons are among the most intense of such crosspeaks in the COSY spectrum recorded at 22 °C (Fig. 1). The lines through the data points were drawn by inspection only. The amide protons of Val 63, Val 118, U8 and U73 are partially accessible to exchange even 100 seconds after initiation of refolding.

Methods. Bovine pancreatic ribonuclease A (Sigma, grade X11A) was purified chromatographically²⁸ and deuterated by dissolving it in D_2O at pH 3, then heating the solution to 60 °C for 20 min; the procedure was repeated after lyophilization. The kinetic experiments were all performed at 5 °C, using a rapid (millisecond) mixing pulsed quench-flow machine that has been described in detail²⁹. The deuterated RNase A was unfolded in an unfolding buffer (2.62 M guanidine hydrochloride, 40 mM glycine, in D_2O , final pH 2); this unfolded RNase A solution (80 mg ml⁻¹) was allowed to equilibrate for 2 h at 5 °C before use. To initiate refolding, the unfolded RNase A solution was diluted 10.5-fold into a refolding buffer (0.442 M ammonium sulphate, 0.11 M sodium borate in D_2O , pH 9.05). At different times after initiation of refolding, the deuterated and partly folded protein solution was diluted 6.25-fold into an exchange buffer (0.4 M ammonium sulphate, 0.25 M guanidine hydrochloride, 0.1 M sodium borate in H_2O , pH 9). The mixing dead-time was 10 ms for both dilution steps. After 10 s, amide hydrogen exchange in the 16% deuterated buffer at pH 9 was quenched by rapidly mixing with a quench buffer (50 mM sodium formate, pH 1.25), so that the final pH was 2.9. The refolding reaction was then allowed to go to completion (10 min) at this pH. For the zero-time point, the refolding buffer (see above) contained H_2O , not D_2O ; thus, refolding and exchange were initiated simultaneously. As for the other time points, exchange was quenched after 10 s, and refolding was allowed to go to completion. The NMR sample of the fully folded RNase A was then prepared as described for the reference spectrum in the Fig. 1 legend. A two-dimensional COSY spectrum of the sample from each time point was recorded at 22 °C. Data recording conditions and data processing were as described for the reference spectrum in Fig. 1 legend. The intensities of the C_α -H-NH crosspeaks (the proton occupancies) in each spectrum were determined by calculating the volume integrals of the crosspeaks, after first setting the baseline of the spectrum to zero. The average error in the determination of the volume integrals was less than $\pm 25\%$. For each spectrum, the average volume of two crosspeaks arising from non-exchangeable protons (the two tyrosine ring-proton crosspeaks, J1 and J2 in Fig. 1) was used as the internal intensity standard. For each C_α -H-NH crosspeak, the intensity of the corresponding crosspeak in the reference spectrum (Fig. 1) was chosen to represent 100% proton occupancy.



Time of refolding (s)

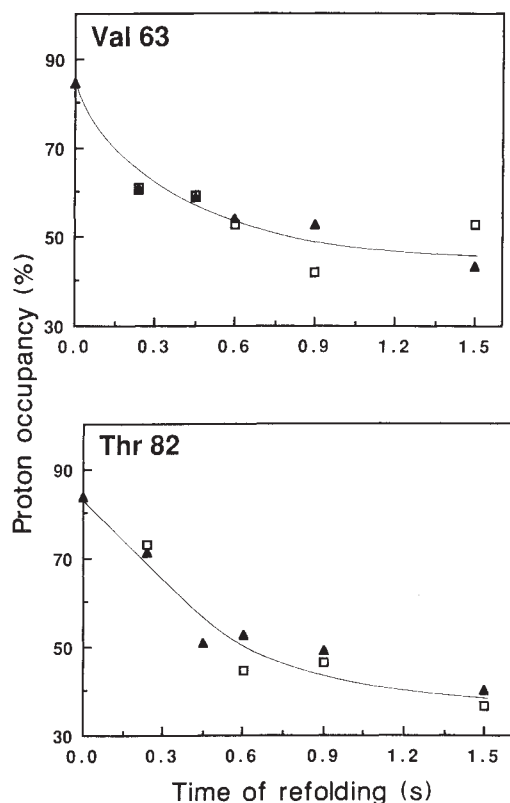


Fig. 3 pH dependence of the formation of the early intermediate. Closed symbols, pH 9; open symbols, pH 10. The experiments at pH 10 were done using the same procedures as those used for the experiments at pH 9 that are described in legends to Figs 1 and 2. Except for the difference in pH, the only other difference was that sodium sulphate was used instead of ammonium sulphate in the buffers at pH 10.

is complete, using ^1H -NMR techniques. It is necessary to obtain ^1H resonance assignments only for the native protein.

Results

In Fig. 2, the accessibility of 8 different backbone amide protons to exchange with solvent protons is plotted against time after initiation of refolding. Deuterated RNase A was allowed to refold in a D_2O buffer for a variable duration of time before exposure to a 10-second proton-labelling pulse at pH 9. Proton-labelling was terminated at the end of the pulse by rapidly dropping the pH of the solution to 2.9. At this pH, amide proton exchange is quenched but folding can proceed to completion. The accessibility to exchange (proton occupancy) of each backbone amide proton site at each time point was determined from the intensity of the corresponding $\text{C}_\alpha\text{H-NH}$ crosspeak in the COSY spectrum of the fully folded protein sample. A control experiment in which native deuterated RNase A was exposed to a 10-second proton-labelling pulse at pH 9, indicated that none of these amide proton sites showed any accessibility to exchange during the exchange pulse (the sites showed zero proton occupancy when assayed by the COSY spectrum). Only 3 out of 40 amide protons showed any degree of proton occupancy in the control experiment, thereby being disqualified from serving as probes for structure formation.

The amide protons shown in Fig. 2, as well as others (data not shown), become protected against exchange according to kinetic curves that are multiphasic and distinct from one another. From the kinetic curves in Fig. 2, it is evident that a kinetic intermediate is formed within the first 1.5 s of refolding. In experiments in which the duration of the proton-labelling pulse, applied 1.5 s after initiation of refolding, was varied between 2

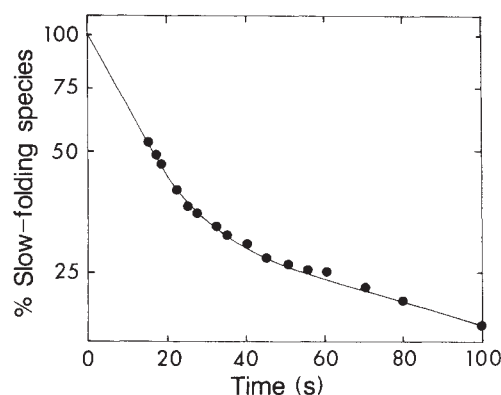


Fig. 4 Refolding kinetics of the slow-folding forms of RNase A. Refolding was monitored by measuring the increase in tyrosine absorbance at 287 nm, which accompanied refolding in a 0.4 M ammonium sulphate, 0.25 M guanidine hydrochloride, 0.1 M sodium borate buffer at pH 9 and 5 °C. The solid line is a double exponential fit through the data; the faster phase has a time constant of 6 s, and the slower phase has a time constant of 128 s. Deuterated RNase A was dissolved in an unfolding buffer at a concentration of 20 mg ml⁻¹; refolding was initiated by a manual 10.5-fold dilution into a refolding buffer. Both unfolding and refolding buffers and consequently the conditions of refolding were the same as those used in the experiments described in Fig. 2.

and 20 s, the extent of labelling at several amide proton sites was unchanged (data not shown). This indicates that the extent of labelling at this time after initiation of refolding does not depend on the exchange behaviour of protected protons in the early folding intermediate. Instead, at least for some amide protons, the extent of labelling depends solely on the fraction of RNase A molecules that have formed the early folding intermediate.

To test this conclusion, the 10-s pulse experiment was repeated at pH 10. Amide proton exchange in intact proteins is base-catalysed above pH 3 in most conditions^{19,20}; thus the driving force for exchange (the intrinsic exchange rate) is most conveniently increased 10-fold by increasing the pH of the proton-labelling pulse from 9 to 10. As seen in Fig. 3, however, the extent of labelling for the Val 63 and Ile 106 amide protons during the first folding reaction detected is not increased at pH 10; several other amide protons also behave similarly (data not shown). This confirms that it is the formation of the early intermediate that is being observed in the pulse-labelling experiments. Under the conditions of the proton-labelling pulse (pH 10, 5 °C), the intrinsic amide proton exchange process in unfolded RNase A has a time constant of 0.6 milliseconds^{19,20}. As the duration of the pulse was 10 s, the exchange rate of each amide proton shown in Fig. 3 must be slowed by a large factor in the intermediate to avoid complete labelling during the pulse (see Discussion). When the pulse is applied 1.5 s after initiation of refolding at pH 9 or 10, the amide proton sites in Figs 2 and 3 are all labelled to an extent of only ~40%, which corresponds to the fraction of protein molecules that have not formed the intermediate during the first 1.5 s of folding. It has been previously shown in the case of the S-protein portion of ribonuclease S (ref. 30), which is a compact globular protein having both secondary and tertiary structure³¹, that the exchange rate constants are reduced only ~1,000-fold (ref. 32). The large reduction in exchange rate constants for some of the amide protons in the folding intermediate is therefore quite remarkable.

The refolding of RNase A can also be monitored by observing the change in tyrosine absorbance (Fig. 4) that accompanies the folding process. Unfolded RNase A consists of a mixture of a fast-folding species, U_F , and at least two slow-folding species, U_S^I and U_S^{II} (refs 1 and 4), which probably arise from proline

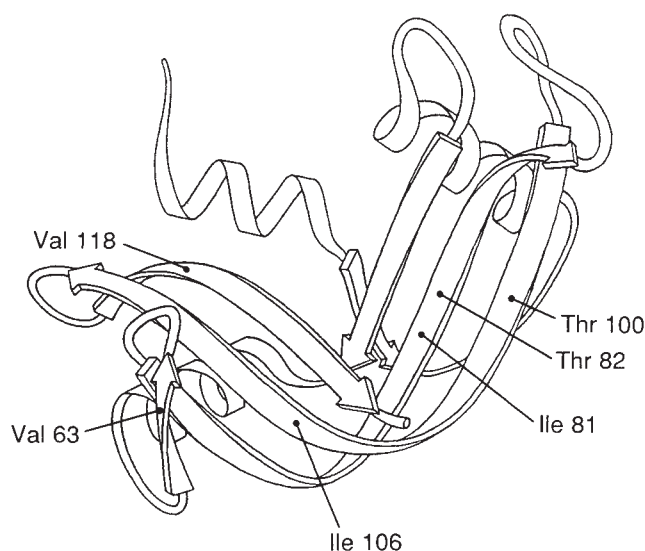


Fig. 5 Tertiary structure of RNase A. The 3 α -helices and the β -sheet are clearly depicted in this ribbon diagram (courtesy of Jane Richardson). The locations of some of the amide protons that have been used as examples in Figs 2 and 3 are shown.

cis-trans isomerization after unfolding⁵. The two kinetic phases observed in the refolding reaction (Fig. 4) reflect the folding of U_S species; the faster phase reflects the folding of the major U_S^H species and the slower phase reflects the folding of the minor U_S^L species⁴. The refolding of U_F is complete within the dead time (10 s) of the manual mixing; in fact, stopped-flow experiments have indicated that the time constant for refolding of U_F is less than 50 ms both at 25 °C (ref. 33) and at 10 °C (ref. 34). It should be noted that the design of the pulse-labelling experiments in Figs 2 and 3 is also such that only the refolding of the U_S species is being monitored; even at the time of application of the earliest pulse, which was 250 ms after initiation of refolding (except for the zero time points), U_F is completely folded³⁴ and therefore will not be labelled.

Tyrosine absorbance monitors the formation of the hydrophobic core of the protein, as solvent is excluded from the vicinity of 3 tyrosine residues that are buried in native RNase A. Like many spectroscopic probes, tyrosine absorbance monitors formation of tertiary structure during folding, whereas the amide proton probes monitor formation of secondary structure, although they may also be sensitive to later formation of tertiary structure. During the folding of both the U_S^L and U_S^H species of RNase A, the tyrosine absorbance probe monitors the formation of a native-like intermediate, I_N (refs 4 and 5), which possesses tertiary structure and subsequently forms native RNase A, probably as the result of one or more proline isomerization reactions⁴. A comparison of Figs 2 and 4 shows that the kinetics of folding of the slow-folding species of RNase A, when measured by the tyrosine absorbance method, are considerably slower than when measured using individual amide protons as probes for folding. This suggests that formation of stable secondary structure precedes the formation of the final tertiary structure during the folding of RNase A, in agreement with earlier low-resolution studies^{2,3}. This result with RNase A is different from the result obtained earlier with bovine pancreas trypsin inhibitor (ref. 35), where the amide proton probes and the tyrosine absorbance probe all showed similar folding kinetics.

A competition method² was used in that study of bovine pancreas trypsin inhibitor³⁵: the competition between folding and amide proton exchange was measured as a function of pH because exchange was base-catalysed, whereas the folding kinetics were nearly independent of pH in the range studied. The experiments reported here demonstrate the power of the pulse-labelling method. In competition experiments, a folding intermediate can be detected only if it is formed rapidly compared with the rate of amide proton exchange from the unfolded

form, which is not the case here. Moreover, if the folding kinetics are complex, as they are for RNase A, the competition experiment does not give information on the kinetics of formation of the intermediate under conditions where it is detected.

Discussion

The kinetics of formation of the early intermediate on the folding pathway of RNase A have been confirmed by [³H]H₂O pulse-labelling experiments (data not shown) using the same rapid (millisecond) mixing method. Earlier manual mixing experiments had shown that an early intermediate, I_1 , with many exchange-resistant amide protons^{2,3}, is formed before the formation of the final tertiary structure. In these low-resolution studies, two mechanisms were considered² for the formation of I_1 from the unfolded protein U . In one mechanism, the equilibrium stability model, I_1 is formed very rapidly from U and a rapid pre-equilibrium is established between I_1 and U ; later, I_1 undergoes the further folding reactions that lead to formation of the final tertiary structure. The mechanism of amide proton exchange from I_1 is transient unfolding of I_1 to U , with exchange into U . I_1 is found to be only moderately stable compared with U according to this mechanism². In a second mechanism, the kinetic model, I_1 is very stable compared with U ; consequently, the formation of I_1 from U is essentially irreversible, and occurs at a rate that is slower than or competitive with exchange into U . The earlier studies could not distinguish between the two mechanisms because neither the stability of I_1 nor the kinetics of formation of I_1 could be directly determined by the methods used, which had both low temporal and structural resolution. The results depicted in Figs 2 and 3 attest to both the high stability of I_1 and a rate of formation of I_1 that, although rapid, is still competitive with the rate of amide proton exchange into U at the pH (~7) which was used in the low-resolution studies^{2,3}. When the rate of amide proton exchange into U is much faster than the rate of formation of I_1 , which is true under the conditions of exchange (pH 9 or 10) used in this study, the second mechanism dictates that the extent of proton-labelling of I_1 should be independent of pH, as we observe (Fig. 3). It thus appears that the second mechanism, the kinetic model, is more appropriate for describing both the results of the earlier low-resolution studies^{2,3} and those reported here.

There are two possible mechanisms^{23,24} to explain the strong and pH-independent resistance (under our experimental conditions) to exchange of the early intermediate I_1 . The first is the EX₂ mechanism, in which a fast pre-equilibrium determines the fraction of I_1 available for exchange. The normal test of this

mechanism is that exchange should be base-catalysed above pH 3. As we do not observe base-catalysed exchange into I_1 (Fig. 3), either exchange does not follow the EX_2 mechanism, or otherwise I_1 is completely resistant to exchange under our conditions. The second mechanism is the EX_1 mechanism, in which the observed exchange rate equals the rate of opening of I_1 to exchange. In this case, the observed exchange rate is not base-catalysed and may be pH -independent. If exchange into I_1 is occurring in our experiments and takes place by an EX_1 mechanism, then the extent of exchange must increase with the duration of the pulse. This will be tested in future experiments. Usually, exchange into intact proteins occurs by the EX_2 mechanism^{23,24}.

In a molecular dynamics simulation of the folding of crambin³⁶ using nuclear Overhauser-effect constraints from NMR data, it has been observed that secondary structure forms before tertiary structure during the folding process. Kinetic circular dichroism studies have indicated that secondary structure is formed before the final tertiary structure during the folding of RNase S (ref. 37), α -lactalbumin^{38,39}, lysozyme³⁹, carbonic anhydrase⁴⁰⁻⁴², cytochrome *c* (ref. 43) and β -lactoglobulin⁴³. Such studies, however, do not indicate whether the units of secondary structure are the same as in the native protein, and little is known about the structural stability of folding intermediates. The low pH form of α -lactalbumin has properties resembling those of the kinetic folding intermediate^{39,44}, and tritium-exchange studies suggest that it has only marginal stability⁴⁵. In contrast the early intermediate, I_1 , on the folding pathway of RNase A has been shown here to possess very stable secondary structure. It has yet to be determined whether or not the simultaneous protection of several β -sheet amide protons in I_1 means that a β -sheet has formed in the intermediate, and if so, whether it has a structure similar to the β -sheet in the native protein. To do this, it will be necessary to obtain kinetic curves of proton occupancy versus time of folding for a much larger number of sequentially assigned amide protons.

Two fundamentally different models for protein folding have been proposed. In the so-called jigsaw puzzle model⁴⁶, which can be traced back to the early two-state model for protein folding, there is neither a defined starting point for folding nor are any unique structural intermediates formed during the folding process. It is proposed that multiple folding pathways exist which are equally accessible to the unfolded protein. Tertiary structure can form without the need for first assembling the backbone; instead, it is assembled in much the same way as a jigsaw puzzle is fitted together. In support of this model, there are numerous possible ways to assemble the native structures of both myoglobin and RNase A by association of compact structural units which are contiguous in the sequences of the

proteins⁴⁷. In the framework model⁷⁻⁹, folding is directed along one or a few well defined sequential pathways with unique structural intermediates. It is proposed that a framework of stable secondary structure is first formed. This secondary structure is presumably stabilized by a small number of specific tertiary interactions; for example, the stabilization of α -helices during the folding of myoglobin is expected to result from as few as 6 principal pairwise helix-helix contacts^{7,8}. The subsequent development of additional tertiary interactions leads to the formation of the native tertiary structure of the protein. Kinetic intermediates on the folding pathway are expected to have at least some stable units of secondary structure. The preliminary results reported here show that the technique used in our studies can distinguish between these two distinct models, and the results support the framework model.

At least 5 of the 8 backbone amide protons that are used as examples in Fig. 2 are located in the β -sheet of native RNase A (Fig. 5). All are protected early on during the folding process. Our preliminary results, although they do not exclude the concurrent formation of other secondary structural units, suggest that the β -sheet may be formed in an early folding reaction of RNase A. It is not possible, at this initial stage of investigation, to interpret the complex kinetics evident at later stages of the folding process (Fig. 2). In part, the complexity must reflect the kinetic heterogeneity of unfolded RNase A (ref. 1); the minor slow-folding species U_S^1 is known to refold last. The existence of more than one folding pathway for any one of the 3 species of unfolded RNase A cannot as yet be ruled out.

In conclusion, our investigations have shown that a kinetic intermediate with many protected amide protons is formed in an early folding reaction of RNase A. The work reported is consistent with the framework model of protein folding, in which formation of stable secondary structure occurs before the assembly of the final tertiary structure. Recently, considerable progress has been made towards a complete sequential NMR assignment of RNase A (A. D. Robertson, E. O. Purisma, & M. A. Eastman, unpublished results). Once this is complete, the use of nearly 40 backbone amide protons as probes for folding will yield a more detailed characterization of structural intermediates formed on the folding pathway.

We thank Kevin Shoemaker for preliminary experiments, Dennis Hare for a copy of the FTNMR program, Virginia Robbins for purifying the protein, and John Osterhout, Andrew Robertson, Heinrich Roder, Franz Schmid, Kevin Shoemaker and David Wemmer for helpful discussions. J.B.U. is a Fellow of The Jane Coffin Childs Memorial Fund for Medical Research. This investigation has been aided by a grant from the Jane Coffin Childs Memorial Fund for Medical Research and by the National Institutes of Health.

Received 13 June; accepted 31 August 1988.

- Garel, J.-R. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3347-3351 (1973).
- Schmid, F. X. & Baldwin, R. L. *J. molec. Biol.* **135**, 199-215 (1979).
- Kim, P. S. & Baldwin, R. L. *Biochemistry* **19**, 6124-6129 (1980).
- Schmid, F. X. *Biochemistry* **22**, 4690-4696 (1983).
- Cook, K. H., Schmid, F. X. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6157-6161 (1979).
- Roder, H., Elove, G. A. & Englander, S. W. *Nature* **335**, 700-704 (1988).
- Pitts, O. B. & Rashin, A. A. *Biophys. Chem.* **3**, 1-20 (1975).
- Richmond, T. J. & Richards, F. M. *J. molec. Biol.* **119**, 537-555 (1978).
- Kim, P. S. & Baldwin, R. L. *Rev. Biochem. Sci.* **51**, 459-489 (1982).
- Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
- Aue, W. P., Bartoldi, E. & Ernst, R. R. *J. chem. Phys.* **64**, 2229-2246 (1976).
- Hahn, U. & Rüterjans, H. *Eur. J. Biochem.* **152**, 481-491 (1985).
- Wand, A. J., Di Stefano, D. S., Feng, Y., Roder, H. & Englander, S. W. *Biochemistry* (in the press).
- Redfield, C. & Dobson, C. M. *Biochemistry* **27**, 122-136 (1988).
- LeMaster, D. M. & Richards, F. M. *Biochemistry* **27**, 142-150 (1988).
- Creighton, T. E. *Prog. Biophys. molec. Biol.* **33**, 231-297 (1978).
- Creighton, T. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5082-5086 (1988).
- States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. *J. molec. Biol.* **195**, 731-739 (1987).
- Kuwajima, K., Kim, P. S. & Baldwin, R. L. *Biopolymers* **22**, 59-67 (1983).
- Kim, P. S. *Meth. Enzym.* **131**, 136-156 (1986).
- Brems, D. N. & Baldwin, R. L. *Biochemistry* **24**, 1689-1693 (1985).
- Roder, H. *Meth. Enzym.* (in the press).
- Englander, S. W. & Kallenbach, N. R. *Q. Rev. Biophys.* **16**, 521-655 (1984).
- Wagner, G. Q. *Rev. Biophys.* **16**, 1-57 (1983).
- Englander, S. W., Donner, N. N. & Teitelbaum, H. A. *Rev. Biochem.* **41**, 903-924 (1972).
- Wlodawer, A. & Sjölin, L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1418-1422 (1982).
- Wlodawer, A. & Sjölin, L. *Biochemistry* **22**, 2720-2728 (1983).
- Garel, J.-R. *Eur. J. Biochem.* **70**, 179-189 (1976).
- Cash, D. J. & Hess, G. P. *Analyt. Biochem.* **112**, 39-51 (1981).
- Richards, F. M. & Vithayathil, P. J. *J. biol. Chem.* **234**, 1459-1465 (1959).
- Richards, F. M. & Wyckoff, H. W. in *The Enzymes 3rd edn* Vol. 4 (ed Boyer, P. D.) 647-806 (Academic, New York 1971).
- Rosa, J. J. & Richards, F. M. *J. molec. Biol.* **145**, 835-851 (1979).
- Garel, J.-R., Nall, B. T. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1853-1857 (1976).
- Lin, L.-N. & Brandts, J. F. *Biochemistry* **22**, 564-573 (1983).
- Roder, H. & Wüthrich, K. *Proteins* **1**, 34-42 (1986).
- Brünger, A. T., Clore, G. M., Gröneborn, A. M. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3801-3805 (1986).
- Labhardt, A. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7674-7678 (1984).
- Gilmanshain, R. I. & Pitts, O. B. *FEBS Lett.* **223**, 327-329 (1987).
- Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. *Biochemistry* **24**, 874-881 (1985).
- McCoy, L. F., Rowe, E. S. & Wong, K. P. *Biochemistry* **19**, 4738-4743 (1980).
- Dolgikh, D. A., Kolomiets, A. P., Bolotina, I. A. & Pitts, O. B. *FEBS Lett.* **165**, 88-92 (1984).
- Sotnikov, G. V. *FEBS Lett.* **224**, 9-13 (1987).
- Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. & Nagamura, T. *FEBS Lett.* **221**, 115-118 (1987).
- Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. *Biochemistry* **25**, 6965-6972 (1986).
- Dolgikh, D. A. *et al. Eur. biophys. J.* **13**, 109-121 (1985).
- Harrison, S. C. & Durbin, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4028-4030 (1985).
- Lesk, A. M. & Rose, G. D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4304-4308 (1981).

Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR

Heinrich Roder, Gülnur A. Elöve & S. Walter Englander

Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059, USA

*To understand the process of protein folding, it will be necessary to obtain detailed structural information on folding intermediates. This difficult problem is being studied by using hydrogen exchange and rapid mixing to label transient structural intermediates, with subsequent analysis of the proton-labelling pattern by two-dimensional nuclear magnetic resonance spectroscopy. Results for cytochrome *c* show that the method provides the spatial and temporal resolution necessary to monitor structure formation at many defined sites along the polypeptide chain on a timescale ranging from milliseconds to minutes.*

A MAJOR obstacle in the ongoing effort to discover the principles of protein folding is the elusive nature of structural intermediates on the folding pathway. The folding transition for globular proteins is often highly cooperative, which makes it difficult to study folding intermediates at equilibrium. Evidence for kinetic intermediates has been found for several proteins¹, but the structural description of these transient states is a formidable experimental challenge. One approach relies on disulphide reactions to trap folding intermediates in stable form (reviewed in ref. 2). We chose a different method which makes use of hydrogen-deuterium exchange to label NH sites on the polypeptide chain that are still unprotected at defined time points during refolding, followed by two-dimensional nuclear magnetic resonance (2-D NMR) analysis of the refolded protein to determine the amount of proton label trapped at points throughout the structure. One can thus follow the time course of structure formation for those regions of the protein where labile hydrogens become protected against exchange on refolding.

The application of hydrogen exchange labelling methods for kinetic folding studies was pioneered by Baldwin and coworkers^{3,4}, who initially used tritium exchange and manual mixing to study the slow folding phase of ribonuclease A. The power of this approach is significantly enhanced by the use of hydrogen-deuterium exchange in conjunction with rapid mixing methods, and the subsequent proton-resolved analysis of the labels trapped during refolding using proton NMR (ref. 5). The feasibility of this method was demonstrated by Roder and Wüthrich⁶ in a study of early structural events in the refolding of bovine pancreatic trypsin inhibitor using one-dimensional NMR. The development of 2-D NMR methods⁷⁻⁹ now makes it possible to monitor a comprehensive set of individual protons at known locations throughout the protein structure. Hydrogen exchange adds the ability to resolve structural events on a millisecond timescale to the inherent spatial resolution of 2-D NMR.

The hydrogen exchange labelling results presented here provide an initial picture of the folding process for cytochrome *c* in terms of the time-resolved formation of hydrogen-bonded structure. Distinct kinetic patterns are observed for amide protons located in different parts of the protein. Several amide sites on the two terminal helices acquire over 60% protection against exchange within the first 20 ms of refolding, a timescale where previously little was known about folding in structural terms. All other amide protons observed so far remain accessible for exchange at the first stage of folding, and are protected by structure formed in two slower kinetic phases on the 100 ms

and 10 s timescale respectively. These include sites in two helical segments between residues 60 and 75, as well as a number of amide protons that form H-bonded tertiary interactions.

The choice of horse cytochrome *c* for these studies was motivated by our success in obtaining essentially complete resonance assignments in the ¹H NMR spectrum in both oxidation states^{10,11,33}. Mitochondrial *c*-type cytochromes are among the best characterized proteins. Several species have been studied by X-ray crystallography¹²⁻¹⁴, and extensive protein-folding studies have been reported¹⁵⁻²¹.

Pulse labelling strategy

Hydrogen exchange labelling experiments with cytochrome *c* require the use of preparative rapid-mixing techniques, because the kinetics of refolding are dominated by processes on the 10 and 100 ms timescale^{16,19}. A variation of the pulse labelling method proposed by Kim and Baldwin⁴ was used for the selective proton labelling of cytochrome *c* (Fig. 1). The protein is initially unfolded in a D₂O-denaturant solution. All exchangeable NH sites become deuterated. Refolding, initiated by rapid dilution of the denaturant, is allowed to proceed for variable time periods before the partially refolded protein is exposed to a 50-ms H₂O labelling pulse. Under the conditions chosen for the pulse (pH 9.3, 10 °C, 40–60 ms), the free-peptide H-exchange time constant is about 1 ms (ref. 22), so that amide sites in still unstructured parts of the protein become fully protonated. On the other hand, the proton label is excluded from sites where exchange is retarded more than 50-fold by prior formation of (H-bonded) structure. The labelling pulse is terminated by a rapid change to slow-exchange conditions and refolding is allowed to proceed to completion. The proton labelling pattern imprinted by the pulse is fixed within the native protein and can be analysed at leisure by 2-D NMR. For the quantitation of individual proton occupancies, we use the well resolved NH-C_αH cross peaks in 2-D *J*-correlated (COSY) spectra^{7,23}.

The requirement for three sequential mixing stages makes it technically difficult to implement the pulse labelling experiment on a rapid-mixing apparatus. The first two mixing events (dilution of the denaturant and dilution of the D₂O solution into H₂O at the labelling pulse) involve large mixing ratios (at least 1:5), which together with the final quenching step results in substantial dilution of the protein (>70-fold). However, 1:1 mixing can be used at the second stage if the D₂O refolding buffer of the original method^{4,24} is replaced by an H₂O buffer under slow H-exchange conditions (dashed line in Fig. 1). This approach is valid as long as H-exchange into the deuterium-

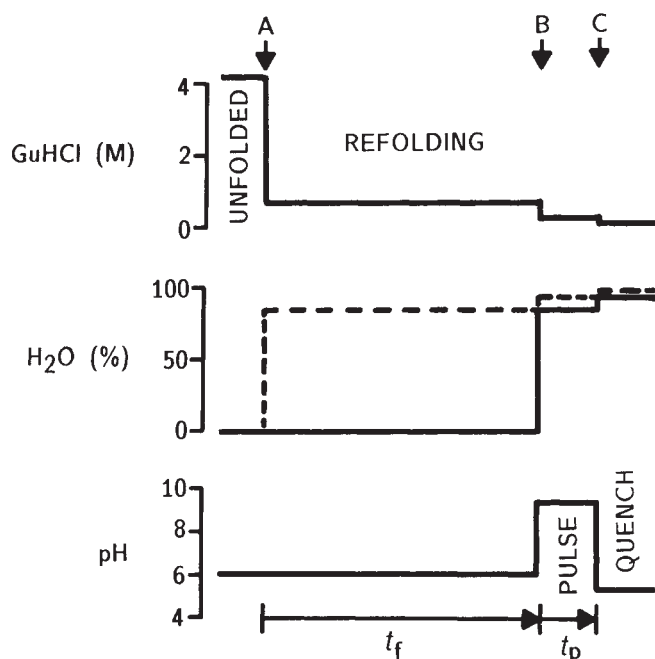


Fig. 1 Schematic illustration of the H-exchange pulse labelling method used in kinetic refolding experiments with cytochrome *c*. The relevant solution conditions are shown in relation to three rapid mixing events (arrows). All experiments were performed at 10 °C on a commercial preparative rapid-mixing apparatus (Hi-Tech, Salisbury, UK) with an estimated mixing dead time of about 2 ms. Oxidized cytochrome *c* (type VI from horse heart, purchased from Sigma) was initially dissolved in D₂O at pD 6.0 containing 4.2 M guanidine hydrochloride (GuHCl) as a denaturant. These conditions were shown to produce complete unfolding of cytochrome *c*, as judged by circular dichroism and hydrogen exchange, with the exception of a limited amount of residual structure in the vicinity of the His 18 haem ligand (unpublished data). Initial protein concentration was 6 mM. Refolding was initiated in the first mixer (A) by rapid 6-fold dilution with 0.1 M acetate at pH 6.2 in H₂O (dashed line). The resulting GuHCl concentration (0.7 M) is well below the unfolding transition. After variable refolding times (t_f), the solution was combined in a second mixer (B) with an equal volume of an H₂O buffer (0.1 M sodium phosphate, 0.1 M glycine) at pH 9.3 (final pH determined in a separate control). Under these pulse conditions, freely accessible amide protons exchange in about 1 ms (ref. 22). The labelling pulse was terminated after 50 ± 10 ms by a rapid pH change to 5.3, accomplished by injecting the solution at the exit of the mixing apparatus into a reservoir of quench buffer (0.3 M sodium citrate, 0.05 M sodium ascorbate, 0 °C) under vigorous stirring (C). The quench conditions favour slow H-exchange (intrinsic exchange times are ~ 10 s) and rapid refolding. The ascorbate in the quench solution serves both as pH buffer and reducing agent for cytochrome *c*, taking advantage of the fact that many amide protons exchange much more slowly in the reduced form of the protein²⁵. This procedure, with H₂O present in the refolding period, was used for experiments at refolding times $t_f < 1$ s. At longer times, H-exchange is no longer negligible (typical free-peptide exchange times are about 3 s at pH 6, 10 °C), and refolding has to be done in D₂O (solid line in centre panel). Subsequent preparation of samples for NMR analysis was carried out at 4 °C. Unfolding was not completely reversible under the conditions used; varying amounts of aggregated or incorrectly refolded protein were detected by NMR. The native monomeric fraction was isolated by use of Biorex-70 cation exchange columns. Washing with 0.7 M NaCl, 0.1 M NaPO₄ (pH 6.0) elutes the native cytochrome *c* and not the incorrectly folded (aggregated) material. The protein was concentrated and transferred into D₂O by repeated dilution and ultrafiltration. The final NMR sample conditions were between 3 and 5 mM reduced cytochrome *c* in D₂O, pD 5.3, 0.05 M K₃PO₄, 0.01 M ascorbate. Samples were stored at 4 °C before NMR measurements performed within 24 hours, or at -85 °C for prolonged periods.

labelled sites of interest is negligible during the refolding time. In the present study this condition was fulfilled for all but the longest refolding times used (> 1 s).

The number of probe points monitored by this method is limited to those hydrogens that exchange slowly in the native protein. In a hydrogen-exchange study of cytochrome *c*, we found that H-D exchange rates can be measured by 2-D NMR methods for about 50 of the 100 backbone amide protons²⁵. Almost all of these protons are involved in crystallographically defined intramolecular hydrogen bonds. The correlation between H-exchange slowing and hydrogen bonding has been noted before²⁶. In most cases then, the protection against hydrogen exchange measured in refolding experiments can be interpreted in terms of hydrogen-bond formation. In the refolding process, some amide hydrogens may become buried and protected against exchange without H-bond formation, but such forms are energetically unfavourable (> 1 kcal mol⁻¹ per proton) and are unlikely to involve many protons. One can also consider the possibility that exchangeable hydrogens may initially become protected by H-bonding to non-native acceptors which are later replaced by the acceptors of the native structure. Such an event might give rise to a transient increase in labelling as t_f (Fig. 1) is varied through the time region where the sites that are protected initially become accessible and are then protected again.

Results

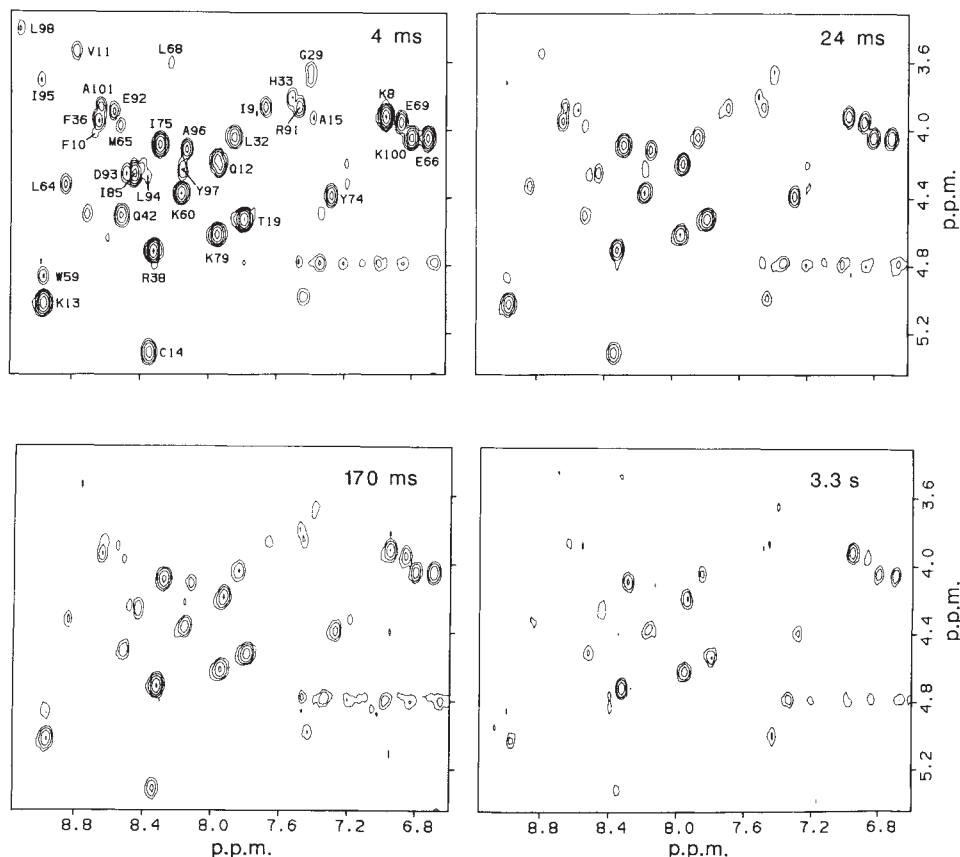
Pulse-labelling experiments with oxidized cytochrome *c* were performed at 12 different refolding times between 4 ms and 1 min. For each time point a separate cytochrome *c* sample was prepared in the reduced form for NMR analysis. A section of the COSY spectrum that includes the majority of NH-C α H cross peaks is shown in Fig. 2 for four representative samples labelled at different refolding times. The extreme upper and lower level of cross-peak intensities (volumes) were determined in control experiments in which either fully unfolded or fully folded cytochrome *c* was exposed to the labelling pulse. The COSY spectrum of the unfolded control sample, labelled to the maximum level determined by the mixing ratio (92%), was used to normalize NH-C α H cross-peak intensities in terms of proton occupancy (intensities vary from one cross peak to another as a result of differences in scalar coupling constants and transverse relaxation times). The folded control experiment was performed to test for the resistance of individual protons in native oxidized cytochrome *c* to H-exchange labelling. Proton label was excluded from all but five amide sites; these are known to exchange rapidly in the oxidized form but slowly in the reduced form²⁵, so they become labelled in the pulse and then trapped when the protein is reduced in the quench buffer. These amide protons therefore cannot serve as structural probes and were disregarded in the analysis.

The extent of labelling at 35 individual amide sites was determined by volume integration of the corresponding NH-C α H peaks, using as an internal intensity standard the average volume of some resolved cross peaks for non-labile protons. For a limited number of amide protons, complementary data were obtained from resolved resonances in the 1-D NMR spectrum. Figure 3 shows the relative proton occupancy as a function of refolding time (logarithmic scale) for a representative set of amide sites. For comparison, Fig. 4 shows the results of a fluorescence-detected stopped-flow refolding experiment under identical conditions. The fluorescence of Trp 59, the single tryptophan residue in horse cytochrome *c*, becomes quenched during refolding as a result of energy transfer to the nearby haem¹⁶.

Kinetic patterns

The pulse labelling results indicate that structural refolding events are dispersed over a timescale ranging from milliseconds to tens of seconds (Fig. 3). Although the detailed behaviour varies among different NH sites, some common kinetic phases can be recognized. The different labelling curves in Fig. 3 can

Fig. 2 Two-dimensional NMR analysis of representative cytochrome *c* samples prepared by the pulse labelling method outlined in Fig. 1. Sections of 2-D J -correlated (COSY) contour plots containing the majority of the $\text{NH}-\text{C}_\alpha\text{H}$ cross peaks are shown for samples labelled at four different refolding times. The residue assignments obtained by Wand *et al.*¹¹ are indicated. Magnitude COSY spectra^{7,23} were recorded at 20 °C on a 500 MHz Bruker AM 500 spectrometer. One hundred and twenty eight transients of 1,024 complex data points covering a spectral width of 9,090 Hz were recorded for each of 400 t_1 increments (0 to 44 ms). The data were processed on a MicroVAX computer, using the FTNMR program provided by Dennis Hare. Unshifted sine multiplication and a 2 Hz line-broadening were applied in both dimensions before Fourier transformation. The final digital resolution was 8.9 Hz per point in both dimensions. Cross-peak intensities were measured by volume integration, using a 3-point radius. The average volume of three cross peaks between non-labile protons (not shown) was used as an internal intensity standard. Relative proton occupancies at individual amide NH sites were obtained by using the unfolded control (see text) for normalization; the COSY cross-peak intensities of the control were taken to correspond to 100% proton occupancy.



be represented as sums of either two or three exponentials. Several sites acquire about 60% protection within about 20 ms, a second common phase occurs at about 200 ms, and all protons become fully protected in a final phase at about 10 s. The small number of distinct kinetic phases suggests that refolding follows a limited number of pathways with discrete intermediates and argues against folding models with numerous parallel folding pathways (compare ref. 27).

At the shortest refolding time used (4 ms), all proton occupancies are reduced by 10–20% relative to the maximum occupancy measured in the unfolded control experiment (Fig. 3). The origin of this effect is uncertain at this time. It may be an indication of an initial rapid folding process occurring in the dead time of the preparative mixing experiments.

Fluorescence observation of the refolding kinetics (Fig. 4) reveals three phases with relaxation times that closely match the main phases of the proton labelling data (Fig. 3). The kinetic trace for Trp 59 fluorescence quenching is similar to the labelling curve for N- and C-terminal protons, but strikingly different from the course of protection of the Trp 59 indole NH (top curve in Fig. 3), which monitors the formation of a tertiary H-bond with the haem propionate. In the initial kinetic phase, the fluorescence signal drops by 40% while the indole NH remains largely protonated, indicating that the tryptophan ring is brought closer to the haem without formation of the specific H-bonded interaction. The discrepancy between fluorescence and proton labelling persists out to the slowest kinetic phase, which has been attributed to kinetic heterogeneity of the unfolded protein¹⁷. Nearly 50% of the molecules form the Trp 59-haem hydrogen bond in the slowest phase, whereas the amplitude of the fluorescence-detected slow phase is only 10%. Such behaviour would clearly not occur in a two-state folding mechanism, even with multiple unfolded forms (see below).

Structural patterns

The degree of protection acquired in the early phase is greatest for amide protons of the N-terminal and C-terminal helices; their proton occupancy drops to about 40% in 30 ms. The examples shown in Fig. 3 are representative for a group of about 10 amide protons, all of which are found in one of the terminal helices of the folded structure. At the same time, other protons located throughout the intervening polypeptide segment remain almost completely accessible for H-exchange labelling; these include sites in two helical segments (residues 61–69 and 71–75), as well as some protons involved in tertiary hydrogen bonds (such as the Trp 59 indole NH to propionate H-bond). Practically all protons analysed so far fall into one of these two classes, although our present data set is not yet complete (quantitation of some amide protons is difficult for technical reasons, for example because of weak $\text{NH}-\text{C}_\alpha\text{H}$ cross peaks or spectral overlap). One possible exception is the backbone NH of His 18, which was found to remain partially protected against H-exchange under denaturing conditions (unpublished observation). It is not surprising to find some residual structure in this part of cytochrome *c*, considering the fact that the polypeptide chain is covalently attached to the haem by Cys 14 and Cys 17, and the His 18 ring remains ligated to the haem iron under the unfolding conditions used.

The rapid protection of N- and C-terminal amide protons in a common kinetic phase distinct from all other protons suggests a special role for the terminal helices at an early stage of refolding. In native cytochrome *c*, the two terminal helices form a tight contact near Gly 6 and Tyr 97, as illustrated in Fig. 5. This interaction is a characteristic feature of all cytochrome *c* proteins of known structure, and the predominantly hydrophobic residues in the contact region are highly conserved. The

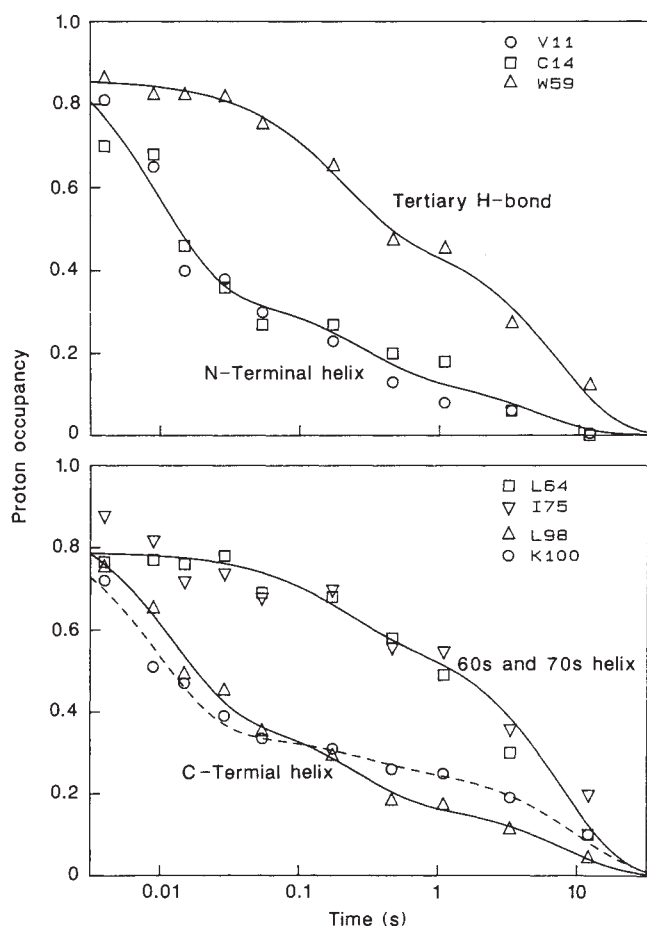


Fig. 3 Time course of protection against H-exchange labelling for a representative set of backbone NH sites, distributed along the cytochrome *c* polypeptide chain, and one side chain NH (Trp 59). Proton occupancies are plotted as a function of the refolding time, t_r , on a logarithmic scale. The curves represent sums of two (top curve in each panel) or three (bottom curve in each panel) exponential terms. Sample preparation and NMR analysis are described in Figs 1 and 2 respectively. As the Leu 98 backbone NH and the Trp 59 indole NH produce weak COSY cross peaks but are well resolved in the 1-D NMR spectrum, these protons were measured by integration of their resonances in 1-D NMR spectra recorded before and after 2-D experiments.

observation that amide protons of both helices are protected on the same rapid timescale suggests that formation of the helices and the helix-helix contact are among the earliest structural events. Although the actual docking of the helices is not directly demonstrated here, it seems unlikely that the two helices would fold independently at the same rate, following the same characteristic labelling curve throughout, while the time course for protection of other segments is very different. The fact that the rate of protection for N- and C-terminal amide protons (50 s^{-1}) coincides with the rate of the initial fluorescence change (reduced Trp 59-haem distance) suggests that association of the chain termini may be accompanied by a general condensation of the initially unfolded chain.

It is noteworthy that in native cytochrome *c* the majority of the very slowly exchanging amide protons are found on the N- and C-terminal helices (Phe 10 and residues 94–99; to be published). The helix contact region of the early folding intermediate is thus among the most stable parts (by H-exchange criteria) of the folded structure. The possibility that the association of chain termini at an early stage of folding may represent a common pattern in protein folding, perhaps as a mechanism for conformational entropy reduction, seems worth considering.

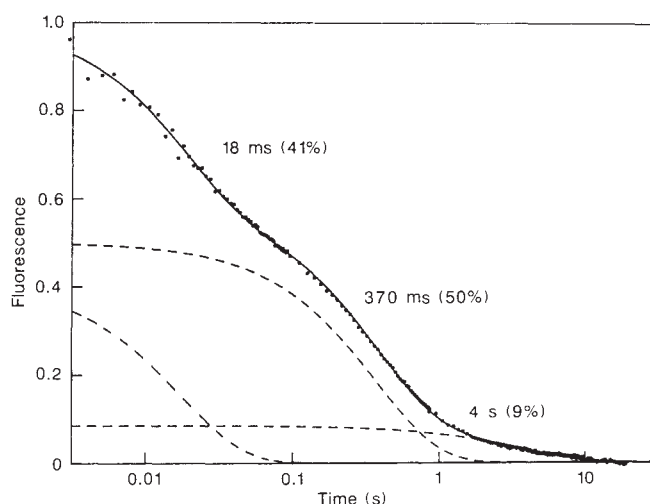


Fig. 4 Refolding kinetics in cytochrome *c* at 10°C , monitored by the fluorescence of a single tryptophan residue (Trp 59). The data were collected on a Hi-Tech stopped-flow apparatus, using a deuterium lamp for excitation at 280 nm and observation of the fluorescence emission at 350 nm. As in the pulse labelling experiments (Fig. 1), cytochrome *c* was unfolded in 4.2 M GuHCl, pH 6.0, and refolded at pH 6.2 in the presence of 0.7 M GuHCl. The final protein concentration was $5 \times 10^{-6} \text{ M}$. Under these conditions, unfolding was fully reversible. The kinetic data are presented on the same logarithmic timescale used in Fig. 3. The solid curve was obtained by nonlinear least-squares fitting of three exponential phases, resulting in the indicated time constants and amplitudes. The individual phases are displayed separately with dashed lines.

In the intermediate kinetic phase ($\sim 200 \text{ ms}$), all observed amide protons acquire at least 50% protection (Fig. 3 and additional data, not shown), indicating the formation of stable structure throughout the molecule. The ultimate level of protection reached in about 1 min (data not shown) is indistinguishable from that of the native protein.

Folding pathways

In terms of folding pathways, let us consider two possible situations. (1) The refolding protein follows a sequential pathway of intermediate states, that is, different structural elements form sequentially and gain stability as refolding progresses. In this case, the time course for protection against H-exchange labelling will differ for different regions of the protein. Furthermore, the degree of H-exchange labelling at each position will depend on the stability of the folding intermediate at that position and the pulse labelling conditions. For example under our conditions, the pulse intensity (pulse time $\times$ free-peptide exchange rate) is about 50, so that an amide site would become 63% labelled if local structure were to reduce its exchange rate 50-fold relative to the free-peptide rate. (2) The unfolded protein consists of a heterogeneous mixture of rapidly and more slowly refolding molecules, and each class of molecules undergoes a cooperative (two-state) folding transition. In this case, varying amounts of unfolded and fully native molecules will be present at different times, but there will be no partially folded forms. All the amide sites in different parts of the protein would then follow the same labelling curve. The degree of labelling at any time point would be as insensitive to the pulse conditions as the native protein.

The present data show some characteristics of both of these extreme possibilities. The observation of distinct labelling curves for different sites provides clear evidence for partially structured folding intermediates. In particular, the rapid protection in the first kinetic phase exclusively for N- and C-terminal amide sites

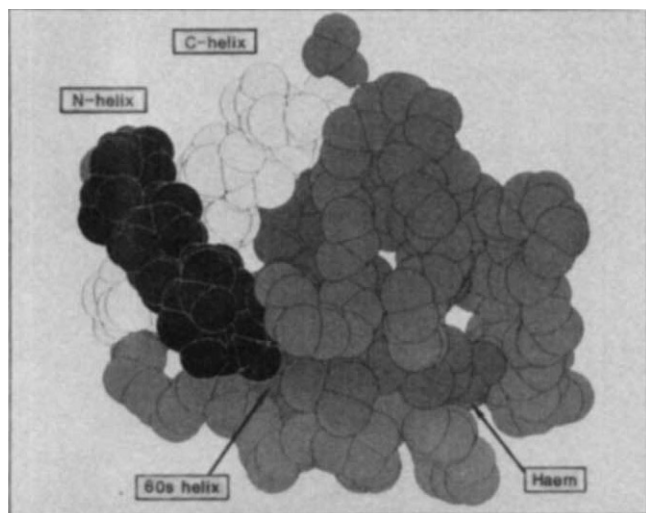


Fig. 5 Space-filling model of cytochrome *c* (ref. 13) showing the location of various structural elements followed in the refolding experiments. Side chains are omitted. The view was chosen to emphasize the contact between the N-terminal (dark) and C-terminal (light) α -helices which is formed early in folding. Also identified are the 60s helix, an example of a structural feature formed late in folding, and the haem. Trp 59 is behind the lower right edge of the haem.

is characteristic of sequential folding behaviour. This result is consistent with the idea that key elements of secondary structure are formed early in folding^{1,5,28}. But the rate-limiting step in stabilizing the early intermediate in cytochrome *c* appears to be the association of the two helical segments, probably driven by hydrophobic interaction of the residues at the helix-helix interface.

The observation of partially folded conformations early in folding argues against the initial model of Brandts and coworkers²⁹, in which a heterogeneous two-state mechanism (the second possibility discussed above) results from the presence of non-native proline isomers that completely block refolding. But the later folding phases are not inconsistent with heterogeneous behaviour. For instance, a possible interpretation for the labelling pattern observed near $t_f = 0.5$ s (Fig. 3) is that about half of the molecules are completely folded, a second fraction (30–40%) contains folded structure only at the chain termini, and a minor population (~15%) of molecules is still extensively disordered, as suggested by the ~15% labelling of

the N- and C-terminal amide sites. In light of the body of data documenting the kinetically heterogeneous nature of unfolded proteins in general^{29–31} and cytochrome *c* in particular¹⁷, some role for the heterogeneous mechanism appears likely. It is interesting to note that the 60s helix (Glu 61–Glu 69) and the 70s helix (Pro 71–Ile 75) are terminated by proline residues (Pro 71 and Pro 76). The presence of *cis* isomer at either of these residues might prevent the proper formation of these helices and produce slow-folding phases.

A definitive answer to how sequential and heterogeneous or parallel mechanisms mix to produce the kinetic refolding pathways(s) will have to await further experimentation. Particularly promising are pulse labelling measurements with variable pulse intensity (pulse duration or pH) which can probe the stability of folding intermediates in terms of their degree of protection against H-exchange.

Discussion

The present study demonstrates the capability of H-exchange labelling methods in combination with 2-D NMR analysis for structural characterization of the transient protein conformations encountered during refolding. The pulse labelling approach applied in this and in the accompanying article⁵ has several advantages compared with the previously used competition method^{3,6,32}: the progress of folding in terms of H-bonded structure formation at individual NH sites can be monitored directly at constant pH; quantitative knowledge of intrinsic H-exchange rates is not required; the interpretation of the data does not rely on kinetic models; and the stability of folding intermediates can be probed by variation of the pulse intensity^{5,24}.

The results can be summarized in terms of a qualitative picture of the cytochrome *c* folding process. In an early kinetic phase, on the 10-ms timescale, only the terminal helices form structure able partially to withstand proton labelling, probably stabilized by native-like helix-helix contacts. This phase is accompanied by some structural condensation, but other features of the native structure, including some helical segments (60s and 70s) and most tertiary H-bond interactions, are still absent. A subsequent kinetic event on the 100-ms timescale, experienced by about 50% of the molecules, is characterized by H-bond formation throughout the molecules affected. At this stage, the remaining slow-folding molecules contain stable structure only in the N- and C-terminal helices. In a final 10-s phase all amide protons acquire the protection against H-exchange labelling characteristic of native cytochrome *c*.

We thank Professors R. L. Baldwin and B. T. Nall for discussions and encouragement. This work was supported by research grants from the National Institutes of Health.

Received 13 June; accepted 5 September 1988.

- Kim, P. S. & Baldwin, R. L. *A. Rev. Biochem.* **51**, 459–489 (1982).
- Creighton, T. E. *Prog. Biophys. molec. Biol.* **33**, 231–297 (1978).
- Schmid, F. X. & Baldwin, R. L. *J. molec. Biol.* **135**, 199–215 (1979).
- Kim, P. S. & Baldwin, R. L. *Biochemistry* **19**, 6124–6129 (1980).
- Udgaonkar, J. B. & Baldwin, R. L. *Nature* **335**, 694–699 (1988).
- Roder, H. & Wüthrich, K. *Proteins* **1**, 34–42 (1986).
- Aue, W. P., Bartholdi, E. & Ernst, R. R. *J. chem. Phys.* **64**, 2229–2246 (1976).
- Ernst, R. R., Bodenhausen, G. & Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon, Oxford, 1986).
- Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
- Wand, A. J. & Englander, S. W. *Biochemistry* **25**, 1100–1106 (1986).
- Wand, A. J., DiStefano, D. L., Feng, Y., Roder, H. & Englander, S. W. *Biochemistry* (in the press).
- Dickerson, R. E. *et al.* *J. biol. Chem.* **246**, 1511–1535 (1971).
- Takano, T. & Dickerson, R. E. *J. molec. Biol.* **153**, 79–115 (1981).
- Louie, G. V., Hutcheon, W. L. B. & Brayer, G. D. *J. molec. Biol.* **199**, 295–314 (1988).
- Ikai, A., Fish, W. & Tanford, C. *J. molec. Biol.* **145**, 265–280 (1973).

- Tsong, T. Y. *Biochemistry* **15**, 5467–5473 (1976).
- Ridge, J. A., Baldwin, R. L. & Labhardt, A. M. *Biochemistry* **20**, 1622–1630 (1981).
- Nall, B. T. & Landers, T. A. *Biochemistry* **20**, 5403–5411 (1981).
- Brems, D. N. & Stellwagen, E. *J. biol. Chem.* **258**, 3655–3660 (1983).
- Ramdas, L. & Nall, B. T. *Biochemistry* **25**, 6959–6964 (1986).
- Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. & Nagamura, T. *FEBS Lett.* **221**, 115–118 (1987).
- Molday, R. S., Englander, S. W. & Kallen, R. G. *Biochemistry* **11**, 150–159 (1972).
- Nagayama, K., Kumar, A., Wüthrich, K. & Ernst, R. R. *J. magn. Reson.* **40**, 321–334 (1980).
- Kim, P. S. *Meth. Enzym.* **131**, 136–156 (1986).
- Wand, A. J., Roder, H. & Englander, S. W. *Biochemistry* **25**, 1107–1114 (1986).
- Englander, S. W. & Kallenbach, N. R. *Q. Rev. Biophys.* **16**, 521–655 (1984).
- Harrison, S. C. & Durbin, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4028–4030 (1985).
- Pitts, O. B. & Rashin, A. A. *Biophys. Chem.* **3**, 1–20 (1975).
- Brandts, J. F., Halvorson, H. R. & Brennan, M. *Biochemistry* **14**, 4953–4963 (1975).
- Garel, J.-R. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3347–3351 (1973).
- Lin, L.-N. & Brandts, J. F. *Biochemistry* **26**, 3537–3543 (1983).
- Roder, H. *Meth. Enzym.* (in the press).
- Feng, Y., Roder, H., Englander, S. W., Wand, A. J. & DiStefano, D. L. *Biochemistry* (in the press).

Counter-rotation in dust-lane ellipticals and the implications for accretion events in galaxies

F. Bertola*, L. M. Buson† & W. W. Zeilinger*

* Dipartimento di Astronomia and † Osservatorio Astronomico, Vicolo dell' Osservatorio 5, I-35122 Padova, Italy

Decoupling of angular momenta of gas and stars in elliptical galaxies constitutes evidence for the acquisition of external material by these galaxies later in their history. In minor-axis dust-lane ellipticals the angular momenta of stellar and gaseous components are orthogonal. Here we present new observations on four major-axis dust-lane ellipticals, Anon 1029-459, NGC3528, NGC4370 and NGC5745. In the first two of these we observe counter-rotation of gas and stars, whereas the two components are co-rotating in NGC4370 and NGC5745; these results are consistent with the acquisition hypothesis. The same accretion events occur also in S0 galaxies, where they give rise to polar rings or to counter- and co-rotating gas in the equatorial plane. As dynamic decoupling in these latter galaxies appears common, we suggest that in all S0s with extended emission lines the gas is of external origin.

The existence of a minor-axis dust-lane in elliptical galaxies¹ has been interpreted as the result of acquisition of a gas-rich galaxy. This idea arises from the fact that the angular momenta of the stellar and the gaseous components are orthogonal²⁻⁷. The existence of major-axis dust-lane ellipticals was later pointed out⁸, in which, by analogy with minor-axis dust-lane ellipticals, it is conceivable that the dust lanes are also of external origin. But it was suggested recently⁹ that this hypothesis requires that an adequate number (50%) of major-axis dust-lane ellipticals must show angular momenta of gas and stars which are antiparallel. Here we present rotation curves of the gaseous component for a sample of four elliptical galaxies which constitute the best-studied objects of this type⁹.

The galaxies Anon 1029-459, NGC3528, NGC4370 and NGC5745 were observed using the Boller and Chivens spectrograph at the Cassegrain focus of the ESO 1.5-m telescope, in the region of $H\alpha$ ($\lambda = 6,300\text{--}7,100 \text{ \AA}$), where dust absorption is lower than at shorter wavelengths. The selected dispersion of the spectra was $59 \text{ \AA mm}^{-1}$ and the scale perpendicular to the dispersion is $87 \text{ arcsec mm}^{-1}$. Observational details are given in Table 1. The observations were carried out by aligning the slit of the spectrograph with the optical major axis in the position of the dust lane. In previous attempts⁹ the $[O II]$ ($\lambda = 3,727\text{--}3,729 \text{ \AA}$) and $[O III]$ ($\lambda = 5,007 \text{ \AA}$) lines were not detected, whereas in the present spectra the $H\alpha$ and $[N II]$ ($\lambda = 6,583 \text{ \AA}$) lines were found in all four objects. Only the stronger of the two lines, namely $H\alpha$ in NGC5745 and $[N II]$ ($\lambda = 6,583 \text{ \AA}$) in all the remaining cases, was measured. The two spectra obtained for Anon 1029-459 were added before measuring. The CCD spectra were calibrated using the standard techniques of dark and bias subtraction, normalized flat-field division and correction for distortion and wavelength calibration. Radial velocities were obtained by applying a gaussian fitting algorithm to the measured line. In none of these spectra are absorption lines detected. The systemic velocities obtained from the emission lines are in good agreement with the values from the absorption lines published previously⁹.

For Anon 1029-459 the $[N II]$ line is sufficiently extended to reach the turn-over point of the rotation curve. For the gas component, we found $V_{\text{max}} = 260 \pm 45 \text{ km s}^{-1}$ at a distance of $R \approx 7 \text{ arcsec}$, giving a gradient $(\Delta V/\Delta R) \approx 40 \text{ km s}^{-1} \text{ arcsec}^{-1}$. For NGC3528 the emission is confined to the inner 6 arcsec

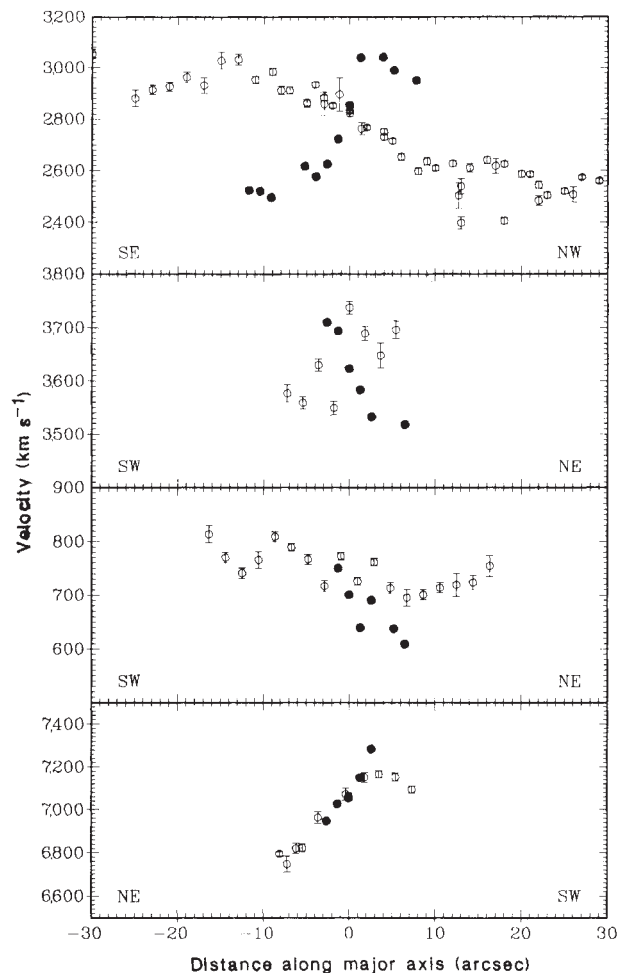


Fig. 1 Velocity curves of the gas (filled circles), compared with that of the stars⁹ (open circles), along the major axis for (top to bottom) Anon 1029-459, NGC3528, NGC4370 and NGC5745.

with a gradient $(\Delta V/\Delta R) \approx 34 \text{ km s}^{-1} \text{ arcsec}^{-1}$. In NGC4370 the emission lines were detected only on the east side, where the continuum is stronger, with a gradient of $(\Delta V/\Delta R) \approx 18 \text{ km s}^{-1} \text{ arcsec}^{-1}$, and for NGC5745 the emission has been detected only in the inner 6-arcsec region, for which we measure a gradient of $(\Delta V/\Delta R) \approx 64 \text{ km s}^{-1} \text{ arcsec}^{-1}$. In the whole sample we find $(\Delta V/\Delta R)_{\text{gas}} > (\Delta V/\Delta R)_{\text{star}}$, which is consistent with the fact that these galaxies are partially pressure supported.

Radial velocities from the emission lines are plotted in Fig. 1, which also includes, for comparison, those from the stellar absorption lines⁹. In NGC4370 and NGC5745 both emission and absorption lines are tilted in the same direction whereas in Anon 1029-459 and NGC3528 the tilts are in the opposite sense. Thus we deduce that in the first two cases the gas and stars are co-rotating but in the latter cases they are counter-rotating. The angular momentum vector for the gaseous component lies in the plane normal to the line of sight and in the direction of the stellar minor axis, as the gas disk is seen edge-on. As far as the stellar body is concerned, in the general case of triaxiality the settling plane of the dust and gas is the one defined by the intermediate and major axes. The apparent minor axis cannot coincide with the intrinsic intermediate axis, because the gas would then occupy the plane defined by the major and minor axes, which is forbidden¹⁰. The strong velocity gradients observed along the major axes and the virtual absence of a velocity gradient along the minor axes⁹ indicate that the stellar component rotates around the intrinsic minor axis. Therefore the angular momentum vectors of the gaseous and stellar

Table 1 Observing log

Object	Date	exp time	Position angle
Anon 1029-459	20-03-1988	150 min	165°
	20-03-1988	180 min	160°
NGC3528	21-03-1988	150 min	60°
NGC4370	21-03-1988	120 min	85°
NGC5745	21-03-1988	90 min	85°

components are parallel in NGC4370 and NGC5745 and anti-parallel in Anon 1029-459 and NGC3528.

Opposite tilts of absorption and emission lines in dust-lane ellipticals have been found previously only in the E0 galaxy NGC5898¹¹, for which observation of counter-rotation, although unambiguous, is less straightforward than in the case observed here. Counter-rotation has also been detected in the elliptical galaxy NGC7097¹², which has no apparent dust lane; but because the stellar component has a very small velocity gradient, the phenomenon is barely observable. In these two cases, as well as in Anon 1029-459 and NGC3528, the counter-rotation is most easily interpreted as an event subsequent to formation of the galaxy, implying acquisition of material from outside. As counter-rotation is expected in 50% of the cases, the dust and gas in the co-rotating cases NGC4370 and NGC5745 should also be interpreted as resulting from accretion. This conclusion is strengthened by the presence of warps in Anon 1029-459⁹, NGC4370 (C. Möllenhoff, private communication) and NGC5745⁹, indicating the presence of material that has not yet settled, as found in minor-axis dust-lane ellipticals⁴. Our observations are consistent with the idea that dust lanes are formed by an acquisition process that involves infall mechanisms of the type described in simulations^{13,14}. As ellipticals are generally triaxial, rotation may occur around any axis and therefore we can envisage alternative angular momentum vector misalignments to those presented here.

Our results help to clarify the nature of accretion events occurring not only in ellipticals but also in S0 and spiral galaxies. Polar-ring S0s and the S0s discovered recently with counter-rotating gas (NGC2217 (G. Galletta, private communication), NGC4546¹⁵ and NGC7007 (Dettmar, private communication)), and the corresponding co-rotating ones, should be considered as the result of similar acquisition processes occurring in minor- and major-axis dust-lane ellipticals respectively. As three out of about a dozen S0s with extended emission lines¹⁶ in the Revised Shapley-Ames Catalogue¹⁷ have already been found to show counter-rotation, and one has a polar ring, we are led to suggest that the gas in these objects is of external origin. The same conclusion has been reached for the H I content of S0 galaxies¹⁸.

Accretion events do not seem to produce pronounced effects in spiral galaxies because of the pre-existing gaseous disk. The gas in the disk will sweep up the polar gas and force it to settle in the plane of the main galaxy¹⁹, although gas infalling into the disk is not likely to produce counter-rotation.

We thank A. Dressler, R. A. E. Fosbury, M. Franx and G. Galletta for many helpful discussions. W.W.Z. acknowledges the support of the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung*, and is affiliated to the Astrophysics Division of the Space Science Department of ESA.

8. Hawarden, T. G., Elson, R. A. W., Longmore, A. J., Tritton, S. B. & Corwin, H. G. Jr *Mon. Not. R. astr. Soc.* **196**, 747-756 (1981).
9. Bertola, F., Galletta, G., Kotanyi, C. & Zeilinger, W. W. *Mon. Not. R. astr. Soc.* (in the press).
10. Steimann-Cameron, T. Y. thesis, Indiana Univ. (1984).
11. Bertola, F. & Bettoni, D. *Astrophys. J.* **329**, 102-106 (1988).
12. Caldwell, N., Kirshner, R. P. & Richstone, D. O. *Astrophys. J.* **305**, 136-147 (1986).
13. Tubbs, A. D. *Astrophys. J.* **241**, 969-980 (1980).
14. Simonson, G. F. thesis, Yale Univ. (1982).
15. Galletta, G. *Astrophys. J.* **318**, 531-535 (1987).
16. Bettoni, D. & Buson, L. M. *Astr. Astrophys. Suppl. Ser.* **67**, 341-351 (1987).
17. Sandage, A. & Tammann, G. A. *A Revised Shapley-Ames Catalogue of Bright Galaxies* (Carnegie, Washington, DC, 1981).
18. Wardle, M. & Knapp, G. R. *Astr. J.* **91**, 23-48 (1986).
19. Schweizer, F., Whitmore, B. C. & Rubin, V. C. *Astr. J.* **88**, 909-925 (1983).

Evidence from mesocosm studies for biological removal of dissolved aluminium from sea water

S. Bradley Moran & Robert M. Moore

Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

The marine geochemical cycle of aluminium is relevant to such oceanographic processes as the diagenesis of aluminosilicate material¹ and tracing atmospheric dust inputs to oceanic surface waters²⁻⁶; Al has also been used as a non-transient water mass tracer⁷. The principal inputs of aluminium to the oceans are from the atmosphere and rivers, and have been reasonably well quantified⁸⁻¹³. Processes controlling its removal from the water column are, however, not well understood and in the past they have been inferred only from measurements of its oceanic distribution^{2-6,12} and unrealistic laboratory culture experiments¹⁴. Here we report the results of a more realistic and controlled temporal study, using a 10-m mesocosm tank, of changes in dissolved aluminium levels and particulate aluminium flux during a diatom bloom. We find that rapid removal of dissolved aluminium occurs by a surface adsorption mechanism and/or incorporation into soft tissues of diatoms. These observations represent the first direct demonstration of the influence and mechanism of biogenic particle removal of aluminium.

Several field studies^{2-6,8} have indicated that the marine distribution of dissolved aluminium is controlled largely by rapid removal processes operating throughout the water column, the result of which is that the most abundant metallic element in the Earth's crust is present in only trace concentrations in the ocean. To understand fully the marine geochemistry of aluminium and thereby to establish its role in various oceanographic processes¹⁻⁷, it is necessary to elucidate the mechanisms controlling its removal from the water column. It is evident that aluminium is very reactive in the marine environment and has a relatively short oceanic residence time, ranging from 35 days in upwelling regions to 50-150 years in deep waters^{4,5}. Vertical and horizontal distributions from the North Atlantic^{6,8} and North Pacific⁴⁻⁶ gyres have been generally explained in terms of a passive particle-adsorption mechanism, but these data provide little direct information on the dynamics of such a process. An apparent Al-Si correlation observed in surface waters of the Mediterranean was interpreted as resulting from aluminium being actively removed by plankton and regenerated along with silicate^{12,15}. Whether aluminium is incorporated into siliceous frustules or is associated with soft tissue has not been satisfactorily resolved. This Al-Si covariance did not hold throughout the water column and has not been observed in open ocean regions. Finally, the unrealistically high concentrations of bio-mass as well as Al (100-900 nmol l⁻¹) used in culture experiments have generally limited their relevance to processes operating in the ocean¹⁴. Consequently, the removal processes that influence the spatial and temporal oceanic distribution of aluminium remain poorly understood.

Received 23 June; accepted 6 September 1988.

1. Bertola, F. & Galletta, G. *Astrophys. J.* **226**, L115-L118 (1978).
2. Sharples, R. M., Carter, D., Hawarden, T. G. & Longmore, A. J. *Mon. Not. R. astr. Soc.* **202**, 37-52 (1983).
3. Caldwell, N. *Astrophys. J.* **278**, 96-111 (1984).
4. Bertola, F., Galletta, G. & Zeilinger, W. W. *Astrophys. J.* **292**, L51-L55 (1985).
5. Wilkinson, A., Sharples, R. M., Fosbury, R. A. E. & Wallace, P. T. *Mon. Not. R. astr. Soc.* **218**, 297-329 (1986).
6. Möllenhoff, C. & Marenbach, G. *Astr. Astrophys.* **154**, 219-226 (1986).
7. Varnas, S., Bertola, F., Galletta, G., Freeman, K. C. & Carter, D. *Astrophys. J.* **313**, 69-88 (1987).

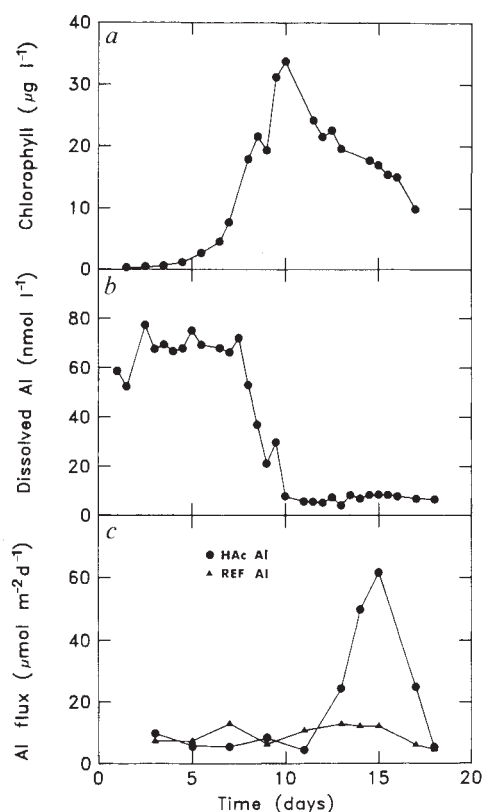


Fig. 1 Temporal variations in *a*, chlorophyll; *b*, dissolved Al; and *c*, fluxes of particulate Al in both acetic acid-leachable and refractory forms (HAc Al (circles) and REF Al (triangles) respectively) observed in the 10-m deep, 117-m³ volume tank. Chlorophyll was measured by fluorescence spectrometry. Surface-adsorbed and soft-tissue particulate aluminium in the biogenic particulate matter collected in the sediment traps was leached with 4.5 M acetic acid⁵. The leachate was stored frozen for later analysis of dissolved aluminium¹⁷ and the residual particulate material remaining on the filter was analysed for aluminium by neutron activation analysis (details will be described elsewhere).

In this work, a mesocosm tank was used to study the response of dissolved aluminium concentration and the vertical flux of particulate aluminium to fluctuations in primary productivity associated with a diatom bloom. The 10-m deep, 117-m³ volume PVC-lined tank was filled with filtered coastal sea water and nutrients (PO₄, Si, NO₃) were added to the illuminated side. The contents were spiked with the common marine diatom *Chaetoceros gracilis*, to seed the bloom, and were illuminated continuously. This experimental approach allows considerable control over the system conditions, which is not possible in the open ocean, but uses ambient aluminium and plankton concentrations. Water samples for measurement of dissolved Al were taken as a time series from a depth of 1 m, filtered, and the filtrate stored frozen for later analysis using the lumogallion method¹⁶. Sedimenting particulate material was collected in plastic sediment traps deployed at 6-m depths for 24-h periods and was then filtered for measurement of acetic acid-leachable (HAc Al) and refractory (REF Al) particulate aluminium. A simultaneous control experiment was run in the dark which showed no significant changes in the nutrients but a slow increase in dissolved aluminium concentration, assumed to be due to contamination; corrections were made for this in the mass balance calculation.

The experiment was characterized by a single pulsed monospecies bloom (Fig. 1*a*). In response to the increased biological activity, the concentration of dissolved aluminium rapidly decreased to near blank levels (Fig. 1*b*), clearly indicating that

Table 1 Experimental variables measured before and after the chlorophyll maximum

Quantity	Initial (day 1)	Final (day 18)
Al (nmol l ⁻¹)	58.8 (59.5)	6.4 (95.8)
Si (µmol l ⁻¹)	14.8 (7.6)	8.6 (10.1)
PO ₄ (µmol l ⁻¹)	7.9 (5.6)	4.1 (4.3)
T (°C)	12 (12)	18 (14)
pH	8.0 (8.0)	8.1 (8.1)
S (‰)	31.16 (31.16)	31.15 (31.16)
Observed loss of Al (nmol l ⁻¹)	—	90
Average Al _{tot} flux (µmol m ⁻² d ⁻¹)*	—	40
Estimated loss of Al (nmol l ⁻¹)	—	85

* Al_{tot} represents total (leachable plus refractory) average particulate aluminium flux. Measurements made at 1 m depth in light and (dark) sides of the mesocosm tank.

algal blooms can cause dramatic decreases in dissolved aluminium. These observations show that phytoplankton growth represents a significant removal process for dissolved aluminium in the marine environment. Our mesocosm experiment has enabled us to examine an important removal process for aluminium in a controlled environment. Such studies are considerably more difficult to pursue in the complex natural environment¹⁷, but nevertheless our results are supported by observations of depletions in dissolved Al in the winter water of the western North Atlantic³. We stress that our results are important to understanding aluminium cycling in the oceans because they represent the first realistic, controlled experimental evidence demonstrating the effect of biological activity on levels and speciation of aluminium.

The approximately sixfold increase in the HAc Al:REF Al ratio of sedimenting material (Fig. 1*c*) suggests that the removal of dissolved aluminium occurs by a surface adsorption mechanism and/or incorporation into soft tissue of diatoms. Active uptake of dissolved aluminium into siliceous frustules of diatoms would be reflected in the refractory particulate aluminium flux but this parameter remained relatively constant for the duration of the experiment. These results are consistent with measurements of opaline aluminium in Antarctic diatom frustules, which were very low and could not be distinguished from terrestrial material in the samples¹⁸. An alternative explanation for this observation might be found in the Al:silicate ratio of these waters, but we are unaware of any dissolved Al measurements for the Southern Ocean.

We can calculate the decrease in dissolved aluminium in the top 2 m of the tank by using an observed average particulate aluminium flux of 40 µmol m⁻² d⁻¹ and by considering that the bulk of the dissolved aluminium was removed over a 7-day period (Fig. 1). The results are summarized in Table 1; there is good agreement between the observed and calculated decrease in dissolved aluminium.

Given that our results show that the intensity of aluminium removal is related to biological productivity, we suggest that these observations may be useful in interpreting and predicting oceanic distributions of aluminium. Seasonal fluctuations in primary productivity, possibly modulated by intermittent mixing events¹⁹, would be expected to alter significantly aluminium levels in surface waters and the resultant fluxes to deep waters²⁰. We note that, because of relatively high biological productivity, coastal and upwelling regions contain lower concentrations of dissolved aluminium than expected. This suggestion is in line with the low aluminium levels reported for the productive coastal surface waters of the eastern North Pacific⁵ and western North Atlantic^{2,21}. The seasonal variations in dissolved aluminium reported in the surface waters surrounding the British Isles²² may also be related to fluctuations in primary productivity, although other processes such as inputs from rivers and sediments may be involved²³. Observations that dissolved aluminium

levels increase with increasing salinity towards open-ocean gyres^{2,5} are consistent with the significant atmospheric contribution of aluminium²⁻⁶ coupled with low biological productivity characteristic of open-ocean regions²⁴.

Aluminium is not considered to be an essential nutritional requirement of marine plankton and is toxic at levels reported for certain acidic lakes and rivers²⁵, but there is some evidence for the existence of aluminium in marine biogenic particulate matter¹⁸. Our results indicate that marine plankton can take up a significant amount of aluminium into a form that is removed by leaching with weak acid. Plausible explanations for this observation include organic binding by cell-surface ligands or active incorporation into cell tissue²⁶. No significant return of particulate Al to solution was observed in our experiment (Fig. 1b), but this could be attributed to its short duration and to the measurements being made in the near-surface waters. The question of regeneration of Al is an important one and requires further study. These results suggest that aluminium could be accumulated along the food chain²⁷.

It is of interest to relate these findings for Al to the behaviour of other trace elements in the ocean. Similar behaviour would be expected for elements that have an analogous speciation in sea water, namely those that exist predominantly as hydrolysed species, $X(OH)_x$ or $X(OH)_y^-$. A close analogue should therefore be Th, which has $Th(OH)_4^0$ as its main dissolved species, as compared with predominant Al species, $Al(OH)_3^0$ and $Al(OH)_4^-$ (ref. 28). In ocean studies using the ^{234}Th - ^{238}U disequilibrium it has been shown that Th is removed from the dissolved state onto algal cells at a rate that varies with the intensity of primary production²⁹. Both elements have short residence times in ocean surface waters; estimates for ^{234}Th and Al in the California Current are 15-50 days and 100 days respectively, and in more productive upwelling areas, 6 days for ^{234}Th and 35 days for Al (refs 5, 29). The longer residence time for Al might be expected because we find that it has a smaller proportion in the colloidal state than does Th (ref. 30). Honeyman and Santschi (manuscript in preparation) have shown that the kinetics of the removal of Th from sea water onto filterable particles are determined essentially by the rate of coagulation of Th-containing colloids. As we observe that Al has a higher truly dissolved fraction, it will not be removed as effectively during the coagulation process.

It should be pointed out, however, that despite these similarities in their removal from ocean waters, ^{234}Th and Al have very different oceanic distributions on account of their different sources and in particular the fact that the atmosphere is an important source only of Al (refs 8-13).

Our results suggest that in general the distributions of those marine trace metals that are influenced by biological activity, such as Cd, Se, As, Cu, Al and Ni, would be expected to exhibit marked spatial and temporal variability. This variability must be considered when interpreting oceanic distributions, calculating residence times and modelling trace-metal behaviour. Because trace-metal distributions reflect several processes operating simultaneously²⁸, such as inputs from the atmosphere and rivers and biological effects, examination of any single control mechanism will be difficult. Although synoptic field surveys provide certain information on the influence of biological activity on trace-metal cycling, this study clearly shows that considerable insight into biological-trace-metal interactions can be obtained using mesocosm experiments. Our results are consistent with the wide variability in the global oceanic distribution of aluminium. They will be useful in the development of a predictive model to describe the complex oceanic behaviour of this marine trace element.

We thank Dr J. Grant and Dr B. Johnson for useful comments on the manuscript. D. Slaunwhite kindly provided chlorophyll measurements; thanks also to Dr X. Zhou, Dr P. Wangersky and R. Pett for providing assistance. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Received 1 June; accepted 30 August 1988.

1. Mackin, J. E. & Aller, R. C. *Mar. Chem.* **14**, 213-232 (1984).
2. Measures, C. I., Grant, B., Khadem, M., Lee, D. S. & Edmond, J. M. *Earth planet. Sci. Lett.* **71**, 1-12 (1984).
3. Measures, C. I., Edmond, J. M. & Jickells, T. D. *Geochim. cosmochim. Acta* **50**, 1423-1429 (1986).
4. Orians, K. J. & Bruland, K. W. *Nature* **316**, 427-429 (1985).
5. Orians, K. J. & Bruland, K. W. *Earth planet. Sci. Lett.* **78**, 397-410 (1986).
6. Hydes, D. J. *Science* **205**, 1260-1262 (1979).
7. Measures, C. I. & Edmond, J. M. *J. geophys. Res.* **93**, 591-595 (1988).
8. Hydes, D. J. *Geochim. cosmochim. Acta* **47**, 967-973 (1983).
9. Chester, R. *Mar. Chem.* **11**, 1-16 (1982).
10. Maring, H. R. & Duce, R. A. *Earth planet. Sci. Lett.* **84**, 381-392 (1987).
11. Uematsu, M. *et al. J. geophys. Res.* **88**, 5343-5352 (1983).
12. Stoffyn, M. & Mackenzie, F. T. *Mar. Chem.* **11**, 105-127 (1982).
13. Hydes, D. J. & Liss, P. S. *Estuar. coast. mar. Sci.* **5**, 755-769 (1977).
14. Stoffyn, M. *Science* **203**, 651-653 (1979).
15. Mackenzie, F. T., Stoffyn, M. & Wollast, R. *Science* **199**, 680-682 (1978).
16. Hydes, D. J. & Liss, P. S. *Analyst* **101**, 922-931 (1976).
17. Moran, S. B. & Moore, R. M. *Estuar. Coast. Shelf Sci.* **27**, 205-215 (1988).
18. Collier, R. & Edmond, J. M. *J. Prog. Oceanogr.* **13**, 113-199 (1984).
19. Wangersky, P. J. *Mar. Chem.* **18**, 269-297 (1986).
20. Deuser, W. G., Ross, E. H. & Anderson, R. F. *Deep Sea Res.* **28**, 495-505 (1981).
21. Moore, R. M. & Millward, G. E. *Geochim. cosmochim. Acta* **48**, 235-241 (1984).
22. Kremling, K. & Hydes, D. J. *Contin. Shelf Res.* **8**, 89-105 (1988).
23. Kremling, K. *Deep Sea Res.* **32**, 531-555 (1985).
24. Parsons, T. R., Takahashi, M. & Hargrave, B. *Biological Oceanographic Processes* 2nd edn (Pergamon, New York, 1977).
25. Schindler, D. W. *Science* **239**, 149-157 (1988).
26. Morel, F. M. M. & Hudson, R. J. M. in *Chemical Processes in Lakes* (ed. Stumm, W.) 251-278 (Wiley, New York, 1985).
27. Eisler, R. *Trace Metal Concentrations in Marine Organisms* (Pergamon, New York, 1981).
28. Bruland, K. W. in *Chemical Oceanography* Vol. 8 (eds Riley, J. P. & Chester, R.) 157-220 (Academic, London, 1983).
29. Coale, K. H. & Bruland, K. W. *Limnol. Oceanogr.* **30**, 22-33 (1985).
30. Niven, S. E. H. & Moore, R. M. in *Radionuclides: a tool for Oceanography* (ed. Guay, J. C.) (Elsevier, Amsterdam, 1988).

Radiocarbon age of last glacial Pacific deep water

N. J. Shackleton*, J.-C. Duplessy†, M. Arnold†, P. Maurice†, M. A. Hall* & J. Cartlidge*

* University of Cambridge, Godwin Laboratory for Quaternary Research, Free School Lane, Cambridge, CB2 3RS, UK

† Centre des Faibles Radioactivités, Laboratoire Mixte CNRS-CEA, 91190 Gif-sur-Yvette, France

One of the most important projects envisaged ever since the possibility emerged of making tandem accelerator mass-spectrometric radiocarbon determinations has been that of investigating changes in the radiocarbon age of the deep ocean¹. Initial attempts to achieve this² were thwarted by difficulties imposed by bioturbation, because these workers used core V28-238 with an accumulation rate of only ~ 1.5 cm kyr⁻¹, thus emphasizing the need to work in a core with a higher accumulation rate. Working in the glacial section of core TR163-31B with an accumulation of 10 cm kyr⁻¹, we show here that the ventilation age of Pacific deep water was about 500 years greater in last glacial times than it is today.

Core Trident 163-31B is a 10×10 cm Kasten core that was taken close to the site of V19-29 and V19-30³ after the importance of these cores became apparent; the coring location is 3°37.2' S, 83°58' W, in a water depth of 3,210 m. About 268 cm of core was recovered. Comparison of the CaCO₃ percentage record in this core with that of the trigger core, as well as that of the trigger cores to V19-29 and V19-30³, shows that ~ 15 cm is missing from the top of the core, probably as a result of difficulties in handling the core on deck after it was retrieved (T. C. Moore, personal communication). In these cores the 10 cm kyr⁻¹ glacial accumulation rate is high enough that bioturbation is no longer a serious concern^{4,5}.

To obtain sufficient material for radiocarbon and other studies, the core was sampled continuously in 1-cm thick segments averaging ~ 10 g dry weight. The sediment is a greenish hemipelagic calcareous mud. Samples were disaggregated by

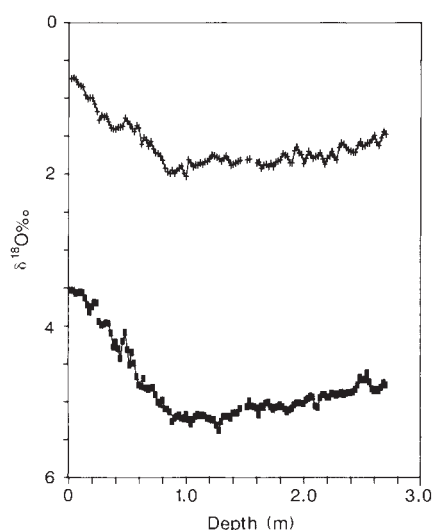


Fig. 1 Oxygen isotope stratigraphy of core TR163-31 (relative to PDB standard). Data are smoothed using a gaussian filter with a total width of 6 cm. ■, benthic (calibrated); +, *N. dutertrei*.

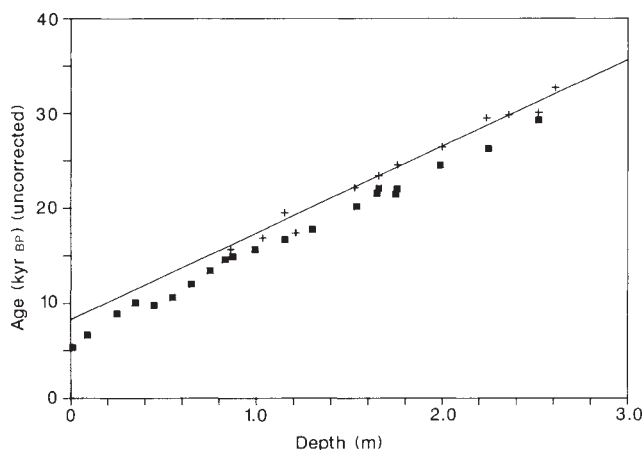


Fig. 2 ^{14}C measurements (5,568 yr half life) for *N. dutertrei* (■) and for *Uvigerina* (+) in core TR163-31 (from Table 1). The regression line (solid) is calculated by weighting the ^{14}C data according to the analytical uncertainties.

gently shaking in distilled water for several hours, and the foraminifera were separated using a 150- μm sieve. Benthic foraminifera were picked from the >250- μm fraction taking particular care to avoid introducing any contaminants. Planktonic *Neogloboquadrina dutertrei* were picked for stable isotope analysis from the 355–425- μm fraction so as to minimize noise deriving from ontogenetic variation in the carbon isotopic composition⁶. To obtain sufficient numbers of *Uvigerina* (~400) for radiocarbon analysis it was necessary to combine specimens from several of the 1-cm levels. The number of specimens taken from each level was recorded, and equivalent mean depth was estimated by weighting each sample depth according to the number of *Uvigerina* specimens contributed. Specimens for radiocarbon analysis were transferred to Gif-sur-Yvette after being cleaned ultrasonically and re-picked under the microscope to eliminate any foreign particles. In Gif-sur-Yvette, approximately the outer 10% was removed in dilute nitric acid immediately before releasing carbon dioxide for conversion to carbon. The stable-isotope methods we used have been outlined in ref. 7, and radiocarbon methods in ref. 8.

Raw stable-isotope measurements are available from NJS; in approximately 150 samples isotopic measurements have been

Table 1 ^{14}C dates

Depth (cm)	<i>dutertrei</i> age* (kyr)	<i>Uvigerina</i> age* (kyr)
today's water	580	2,100
1	5,340 $\pm$ 120 (1–2)	
9	6,660 $\pm$ 110 (9–10)	
25	8,900 $\pm$ 130 (25–26)	
35	10,050 $\pm$ 180 (35–36)	
45	9,800 $\pm$ 150 (45–46)	
55	10,640 $\pm$ 160 (55–56)	
65	12,020 $\pm$ 170 (65–66)	
75	13,440 $\pm$ 250 (75–76)	
83	14,600 $\pm$ 180 (83–84)	
85		15,660 $\pm$ 270 (81–91)
87	14,920 $\pm$ 220 (87–88)	
99	15,660 $\pm$ 200 (99–100)	
103		16,850 $\pm$ 230 (101–108)
114	16,700 $\pm$ 260 (114–115)	19,510 $\pm$ 330 (113–118)
121		17,400 $\pm$ 240 (119–126)
130	17,780 $\pm$ 220 (130–131)	
153		22,140 $\pm$ 310 (151–155)
154	20,160 $\pm$ 300 (154–155)	
165	21,540 $\pm$ 370 (165–166)	
166	22,060 $\pm$ 280 (166–167)	23,420 $\pm$ 310 (165–168)
175	21,470 $\pm$ 360 (175–176)	
176	22,030 $\pm$ 250 (176–177)	24,530 $\pm$ 470 (175–178)
199	24,510 $\pm$ 320 (199–200)	26,430 $\pm$ 410 (199–201)
224		29,530 $\pm$ 550 (223–226)
225	26,270 $\pm$ 460 (225–226)	
235		29,870 $\pm$ 590 (235–237)
251		30,080 $\pm$ 660 (251–253)
252	29,270 $\pm$ 600 (252–253)	
261		32,700 $\pm$ 910 (259–263)

Half life of ^{14}C is 5,568 yr. No reservoir corrections to data.

* Numbers in parentheses indicate depth range in cm of sample.

made of *Uvigerina senticosa*, *Cibicidoides wuellerstorfi* and *N. dutertrei*, at about 2-cm intervals. In Fig. 1 the data are plotted against depth in the core, after smoothing in the depth domain with a gaussian filter with total width of 6 cm. The purpose of this is to reduce the effect of analytical measurement noise ($\pm 0.07\text{‰}$ for a single measurement) without adding any smoothing that exceeds what must already have been imposed naturally by bioturbation. The scale of oxygen isotopic variation in both benthics and planktonics is very similar to that observed in core V19–30⁹. We determined the abundance of *N. dutertrei* and of *U. senticosa* at ~5-cm intervals; below ~70 cm there is no trend in abundance of either and ^{14}C dating is straightforward³. In the upper 70 cm there are marked changes in species abundance, whereas intense dissolution is found in the upper part of the trigger core; interpreting ^{14}C dates in the upper 70 cm will be therefore more complicated³. Our objective here is to investigate the ^{14}C 'age' of deep water in the glacial period only.

^{14}C determinations are shown in Table 1, expressed in radiocarbon years (5,568 yr half life) without any adjustment being made for the 'age' of sea water. Recent (pre-bomb) surface water in the equatorial East Pacific has an 'age' of ~580 yr (ref. 10), whereas deep water has an 'age' of ~2,100 yr at nearby GOSECS Station 331¹¹. The raw ages for *N. dutertrei* and *Uvigerina* are plotted against depth in Fig. 2. If the ventilation of the deep Pacific had been unchanged, the two data sets would show a constant offset of ~1,520 yr; in fact, below ~1.3 m (that is, earlier than 17 kyr) the mean difference is ~2,000 yr. This can arise only if the exchange of CO_2 with the deep ocean was less efficient in glacial times than today. Another recent study has yielded a similar estimate for the ventilation of the deep Pacific¹².

Under the usual convention for radiocarbon ages of marine materials, it is assumed that the radiocarbon age of the surface in the region has not changed substantially. For the purpose of plotting data on a timescale, one would use the radiocarbon data from planktonic species as these ages would be compared

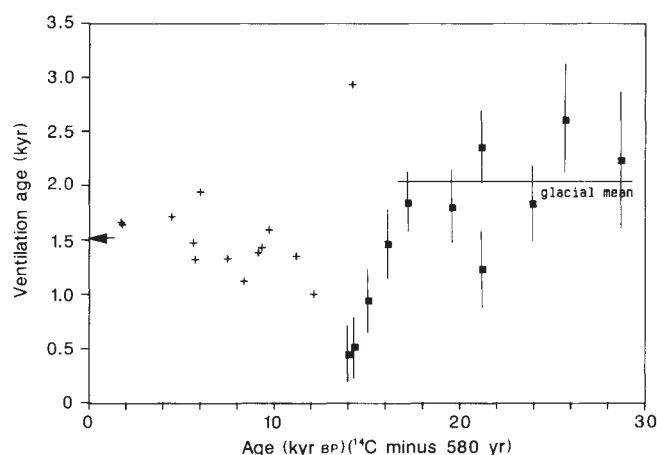
Table 2 Estimates of ventilation age of Pacific deep water, based on data in Table 1

Depth (cm)	Regression benthic age (kyr)	Measured planktonic age (kyr)	Difference (kyr)	Atmosphere ^{14}C age (kyr)
(today)	2,100	580	1,520	0
83	15,051	14,600	451	14,020
87	15,440	14,920	520	14,340
99	16,610	15,660	950	15,080
115	18,168	16,700	1,468	16,120
130	19,629	17,780	1,849	17,200
154	21,967	20,160	1,807	19,580
165	23,038	21,780	1,258	21,220
176	24,109	21,750	2,359	21,170
199	26,350	24,510	1,840	23,930
225	28,882	26,270	2,612	25,690
252	31,512	29,270	2,242	28,690

more readily with those derived from terrestrial organic materials. But the data in Fig. 2 and Table 2 show that the difference in radiocarbon age between surface and deep water was greater during glacial time than today by ~ 500 yr; we will argue that this has a major implication for timescale development.

Normally we use ^{14}C for dating on the assumption that the ^{14}C content of the atmosphere has remained constant. Exceptionally, for the past 7,000 years we are able to use 'calibrated' dates that are based on measurements of the ^{14}C content of dendrochronologically dated segments of long-lived trees¹³. But as there is much more carbon in the ocean than in the atmosphere, changes in the circulation rate (or more precisely, the carbon dioxide ventilation rate) of the deep ocean will affect the ^{14}C content of the atmosphere. Indeed one of the motives for making the measurements presented here is to gain insight into the past ^{14}C activity of the atmosphere. If the production of ^{14}C is assumed to have been constant but the rate of ocean carbon ventilation is variable (as our data show must be the case), the quantity that is more nearly proportional to true age is not the ^{14}C content of plant material photosynthesized from the atmosphere, but the ^{14}C content of materials that derived their carbon from the deep ocean. The standard for zero age would then not be modern wood, but modern deep-ocean dissolved carbon; to a first approximation we can estimate this by subtracting 2,100 from the measured 'age' of benthic foraminifera in core TR163-31B.

Having thus established that the ^{14}C age of the deep ocean has indeed changed relative to that of the surface during the past 30 kyr, it is appropriate to develop an age model for the core using the benthic rather than the planktonic ^{14}C data to investigate the history of ventilation age. On Fig. 2 a regression line is shown through the ^{14}C data for *Uvigerina*. Between ~ 80 cm and the bottom of the core there is no evidence for any departure from this line, implying a steady accumulation of ~ 10.6 cm kyr⁻¹. Figure 3 shows the history of the ^{14}C gradient between surface and deep water, as estimated by deviations of the measurements of ^{14}C in *N. dutertrei* from this regression line (for Fig. 3 the pair of measurements at 175 and 176 cm have been averaged). Also shown on Fig. 3 are the estimates of deep-water ventilation age by Andree *et al.*¹⁴. For convenience of comparison with other data, the ventilation age is plotted as a function of conventional 'atmospheric' ^{14}C age (that is, after subtracting 580 yr from the age measured in *N. dutertrei*). It is clear that before ~ 17 kyr BP there is a significant positive deviation from today's value; the seven earliest estimates have a mean of $1,995 \pm 450$. This implies that the atmospheric ^{14}C content must have been higher than today by an amount equivalent to a dating difference of -500 yr (that is, ^{14}C content of $\sim 60\%$). Between 17,000 and 14,000 yr ago this situation was

**Fig. 3** Pacific ^{14}C ventilation age estimates from core TR163-31 (■) (Table 2) and from Andree *et al.*¹⁴ (V35-5, V35-6; +). The arrow indicates the value for today's ocean.

apparently reversed so that by 13,500 yr ago the atmosphere was lower in ^{14}C than it is today ($^{14}\text{C} \sim -50\%$). Thus after significantly reduced ventilation through late stage 3 and most of stage 2, intense renewed ventilation began around 17 kyr BP. If we assume for discussion that this renewed ventilation began exactly 17,000 yr ago, 17,000-yr-old plant material would have a ^{14}C age of $\sim 16,500$ BP, while material 14,000 yr old would have a ^{14}C age of $\sim 14,500$ BP, solely on account of changes in ventilation rate. Thus although the ventilation age differences are not dramatic, they are sufficient to have a major impact on estimated rates of change around the beginning of deglaciation. Although we do not yet have good control on the precise timing of the onset of renewed ventilation, the combination of our data and that of Andree *et al.*¹⁴ shows that it must have been very close to the beginning of deglaciation—just when, on intuitive grounds, one would expect the opposite on the basis that the ocean should have begun to stabilize as a result of meltwater influx.

Areas like the East Pacific were initially rejected for this study¹ because upwelling makes the surface waters somewhat depleted in ^{14}C relative to other open ocean regions, resulting in a ^{14}C 'age' at the surface of 580 yr (ref. 1) rather than the open ocean value of 400 yr (ref. 15). It is obvious that this complication cannot significantly affect our conclusions, however, because to increase the age difference between *N. dutertrei* and *Uvigerina* by appealing to surface changes one would require an enormous increase in the residence time of water at the surface during glacial times, but all the evidence points to increased productivity¹⁶ in this region, indicating, if anything, a decreased surface residence time.

Thus, the ^{14}C ventilation age of the deep Pacific around glacial maximum time (30–17 kyr BP) was $(1,995 \pm 450)$ yr, 500 yr greater than it is today; at the beginning of deglaciation ventilation of the deep ocean improved rapidly, reducing the ventilation age to ~ 500 yr less than today. In conventional ^{14}C terms, the ^{14}C activity of the atmosphere dropped by $\sim 100\%$ within 3,000 yr or less, perceptibly compressing the radiocarbon timescale of events in this critical interval of time.

We acknowledge support from NERC and EEC. We thank M. Leinen, A. McIntyre, T. Pedersen, R. Grün, J. Le and R. Corfield for assistance.

Received 6 June; accepted 1 September 1988.

1. Broecker, W. S., Mix, A., Andree, M. & Oeschger, H. *Nucl. Instrum. Meth. Phys. Res.* **B5**, 331–339 (1984).
2. Andree, M. *et al. Geophys. Monogr.* **32**, 143–153 (1985).
3. Ninkovich, C. D. & Shackleton, N. J. *Earth planet. Sci. Lett.* **27**, 20–34 (1975).
4. Peng, T.-H. & Broecker, W. S. *Nucl. Instrum. Meth. Phys. Res.* **B5**, 346–352 (1984).

5. Bard, E., Arnold, M., Duprat, J., Moyes, J. & Duplessy, J.-C. *Clim. Dynamics* **1**, 101-112 (1987).
6. Berger, W. H., Killingley, J. S. & Vincent, E. *Oceanologica Acta* **1**, 203-261 (1978).
7. Shackleton, N. J., Hall, M. A. & Boersma, A. *Init. Rep. DSDP* **74**, 599-612 (1984).
8. Arnold, M., Bard, E., Maurice, P. & Duplessy, J.-C. *Nucl. Instrum. Meth. Phys. Res.* **B29**, 120-123 (1987).
9. Shackleton, N. J., Hall, M. A., Line, J. & Cang, S. *Nature* **306**, 319-322 (1983).
10. Druffel, E. M. *Geophys. Res. Lett.* **8**, 59-62 (1981).
11. Ostlund, H. G. & Stuiver, M. *Radiocarbon* **22**, 25-53 (1980).
12. Broecker, W. S. *et al. Paleoceanography* (in the press).
13. Damon, P. E. in *Nuclear and Chemical Dating Techniques*, ACS Symp. Ser. (ed. Curry, L. A.) 233-244 (Am. Chem. Soc., Washington, DC, 1982).
14. Andree, M. *et al. Clim. Dynamics*, **1**, 53-62 (1986).
15. Broecker, W. S. & Peng, T.-L. *Tracers in the Sea* 236-274 (Eldigio, New York, 1982).
16. Pedersen, T. F. *Geology* **11**, 16-19 (1983).

Seismic delineation of upthrust Archaean crust in Kapuskasing, Northern Ontario

A. V. Boland*, R. M. Ellis*, D. J. Northey†, G. F. West†, A. G. Green‡, D. A. Forsyth‡, R. F. Mereu§, R. P. Meyer||, P. Morel-à-l'Huissier‡, G. G. R. Buchbinder‡, I. Asudeh‡ & R. A. W. Haddon‡

* Department of Geophysics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada

† Geophysics Laboratory, Physics Department, University of Toronto, Toronto, Ontario M5S 1A7, Canada

‡ Geological Survey of Canada, Department of Energy, Mines and Resources, Ottawa, Ontario K1A 0Y3, Canada

§ Department of Geophysics, University of Western Ontario, London, Ontario N6A 5B7, Canada

|| Department of Geology and Geophysics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

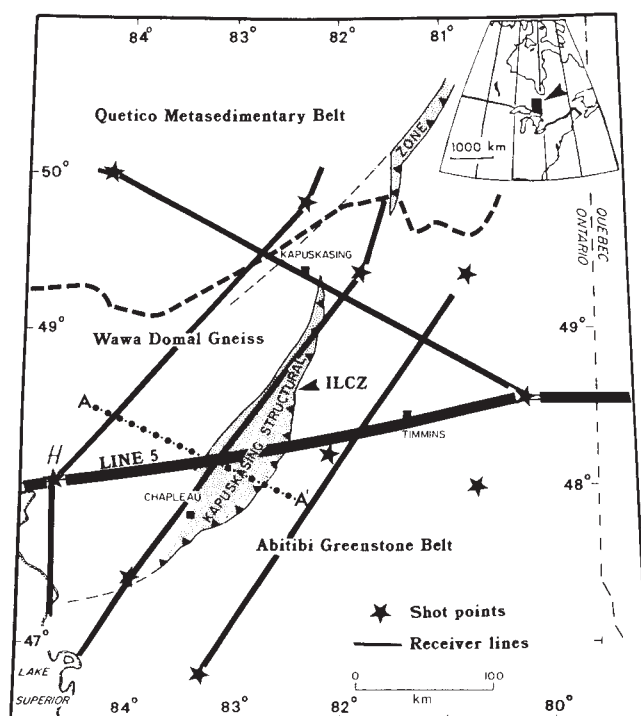


Fig. 1 Locations of shot points and receiver lines in the 1984 seismic refraction experiment. Main geological terranes are shown along with the location of the Ivanhoe Lake cataclastic zone (ILCZ). Inset shows the location of the study area in western North America. The heavy solid line is the profile modelled here; A-A' is the schematic geological cross-section of Fig. 4a. The KSZ is shaded.

The Kapuskasing structure (Fig. 1) is a 500-km-long, high-grade metamorphic zone which cuts obliquely across the predominantly east-west structural grain of the Superior Province of the Canadian shield. Results from surface geological mapping and gravity profile analysis suggest that the structure is a partial cross-section of Archaean crust exposed by thrust faulting along the Ivanhoe Lake cataclastic zone during the early Proterozoic^{1,2}. Similar exposures of the crust have been observed elsewhere^{3,4}, but no seismic investigation has been conducted over a contiguous, exposed cross-section of Archaean crust. Here we report results from a seismic refraction profile across the Kapuskasing structural zone, carried out to investigate the deep structure of this greenstone-gneiss belt. Modelling of travel-times and amplitudes yields a high-velocity anomaly of 6.6-6.7 km s⁻¹ extending from 20 km depth to close to the surface under the Kapuskasing zone, with a suggested dip of 15 ± 2° to the west. This anomaly is in agreement with a structure comprising upthrusting middle to lower crust.

The Michipicoten belt, to the west of the map region, and the Abitibi belt in the east consist of metavolcanic and meta-sedimentary suites of similar ages and composition⁵. Homogeneous tonalitic gneisses and granodiorites characterize the intervening Wawa domal gneiss terrane and layered mafic gneiss, paragneiss and anorthosite outcrop in the Kapuskasing structure itself. There is a transition in metamorphic grade from low-grade greenschist facies in the Michipicoten belt through amphibolite facies in the Wawa domal gneiss terrane to high-grade upper amphibolite and granulite facies in the Kapuskasing structure (Fig. 1). The Ivanhoe Lake cataclastic zone (ILCZ) marks an abrupt return to lower-grade greenschist and amphibolite facies in the Abitibi belt to the east¹. The west-east transition is characterized by an increase in metamorphic formation pressure from 2-3 kbar greenschists through 4-6 kbar amphibolites to 7-9 kbar granulites⁶. The observed amphibolite/granulite transition coincides with an increase in average density⁷ from 2.7 to 2.82 g cm⁻³ and an increase in seismic velocity, at 600 MPa confining pressure, from 6.55 to

6.90 km s⁻¹ (D. M. Fountain and M. H. Salisbury, personal communication). This transition may correspond, therefore, to the Conrad or other similar reflective zones, as imaged seismically at depth in the Superior Province⁸⁻¹⁰. The above results and observations, combined with gravity profile modelling^{1,6} across the structure, suggest that the Kapuskasing structural zone (KSZ) is an uplifted oblique cross-section through the upper two-thirds of the Archaean crust^{1,11}. Thus, the seismic structural analysis presented here provides key information at depth about this Precambrian structure.

In July 1984, as part of the multidisciplinary Lithoprobe program¹², five seismic refraction lines of 360-450-km length were shot over the area. Eighteen profile shots and two fan shots were made, with a recorder spacing of 2-5 km (Fig. 1). The travel-times and amplitudes of the seismic data were modelled using an algorithm¹³ based on zero-order asymptotic ray theory¹⁴. Data from both first and secondary arrivals were used in this procedure. The first arrivals were used to obtain a velocity structure; the intracrustal and Moho wide-angle reflections were used to obtain greater structural definition.

Modelling was refined for all five lines until satisfactory agreement was obtained between the observed and calculated seismograms and the models were matched at the line intersections. Here we discuss the key results from line 5, which was reversed and centre-shot. The synthetic seismogram modelling procedure is illustrated in Figs 2 and 3 for shot H of line 5, which lies across the southern part of the KSZ and provides a cross-sectional view of the structure. Figure 2b shows a simplified ray diagram, used to produce the travel-times and amplitudes. The ray model comprises a number of sub-horizontal layers, each of which is broken up into blocks that have fixed velocities at the top and bottom. Reflections from these layer boundaries can be generated and have been included, when necessary, to model the later arrivals. We do not assume that laterally continuous reflecting horizons exist in the crust, but

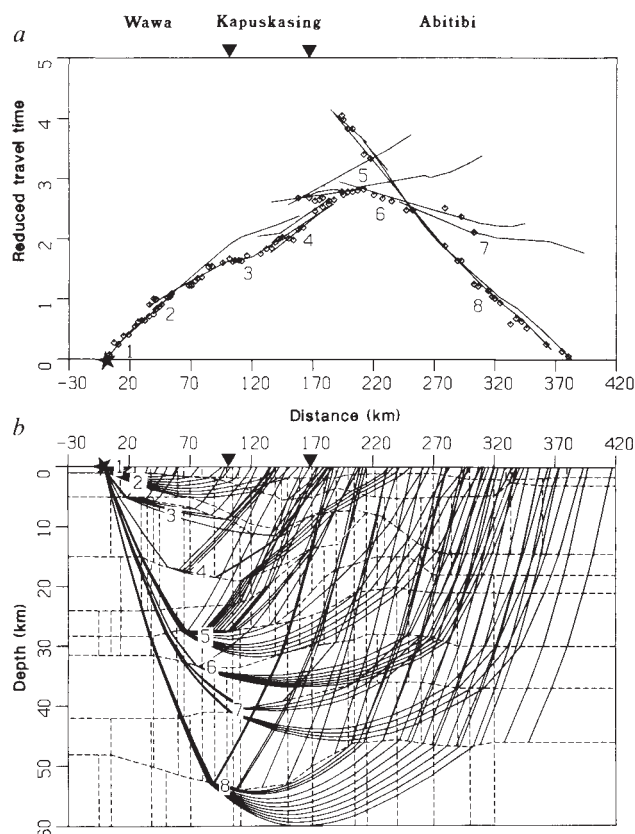


Fig. 2 An example of the ray modelling procedure used in the analysis of the data from shot H of line 5. *a*, The observed reduced travel-times (symbols) compared with those calculated from paths through a block model in a ray-tracing procedure (solid curves). Reduced travel-time = time - (distance/7 km s⁻¹). *b*, Simplified ray model which, for clarity, uses only a few rays in each refracted family and shows the ray groups associated with each of the numbered travel-time branches of *a*. Boundaries of the geological terranes are indicated by arrowheads on each diagram.

where wide-angle reflections are observed the short-imaged reflective zone is included in the modelling procedure as a reflecting boundary. Upper-mantle refractions are modelled by turning rays in the upper mantle and critical refractions from the Moho. Numbered ray groups in Fig. 2*b* correspond to the appropriate travel-time curves in Fig. 2*a*. The latter shows a comparison of the calculated travel-time curves (solid lines) with the picked data (symbols) for the refracted arrivals. Errors in picking and modelling result in uncertainties of ± 0.05 km s⁻¹ in the velocities and of ± 0.5 km in the boundary depths. In view of the perturbing effects of a realistic heterogeneous crust on travel-times and amplitudes¹⁵, however, more realistic uncertainties in the upper and middle crust are ± 0.1 km s⁻¹ and ± 1.0 km respectively. The seismic section exhibits high-frequency back-scattered energy with a few correlatable secondary features (Fig. 3*a*). The section is characterized by these high-amplitude later events and by generally weak first arrivals. There is a noticeable travel-time pull-down of the primary low-amplitude events at 110–170 km over the KSZ. Modelling of these arrivals identifies an anomalous high-velocity zone in the upper crust (Fig. 2). The unusual high-frequency energy, attenuated by scattering, fades near 160–180 km offset, after the Ivanhoe Lake cataclastic zone has been crossed, so that the Moho reflection is diffuse and difficult to trace and the upper-mantle arrivals are weak and often buried in noise. The synthetic section (Fig. 3*b*) was constructed from the amplitudes associated with the calculated travel-times of Fig. 2*a*. Two wavelets were extracted from the real data and convolved with the amplitude spikes to obtain synthetic seismograms similar in spectral character to

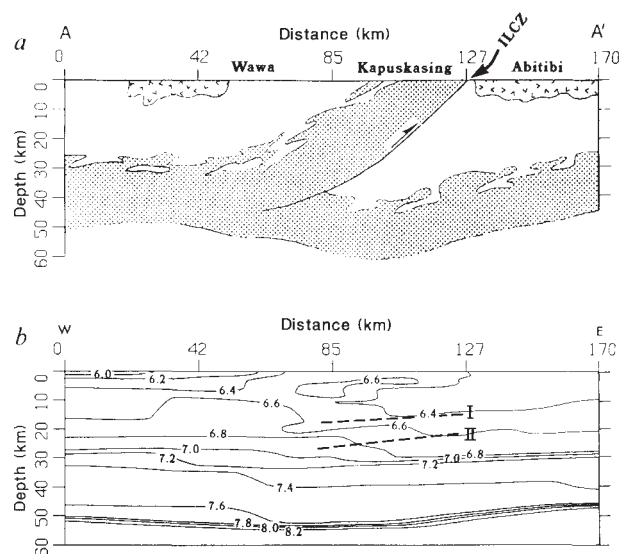


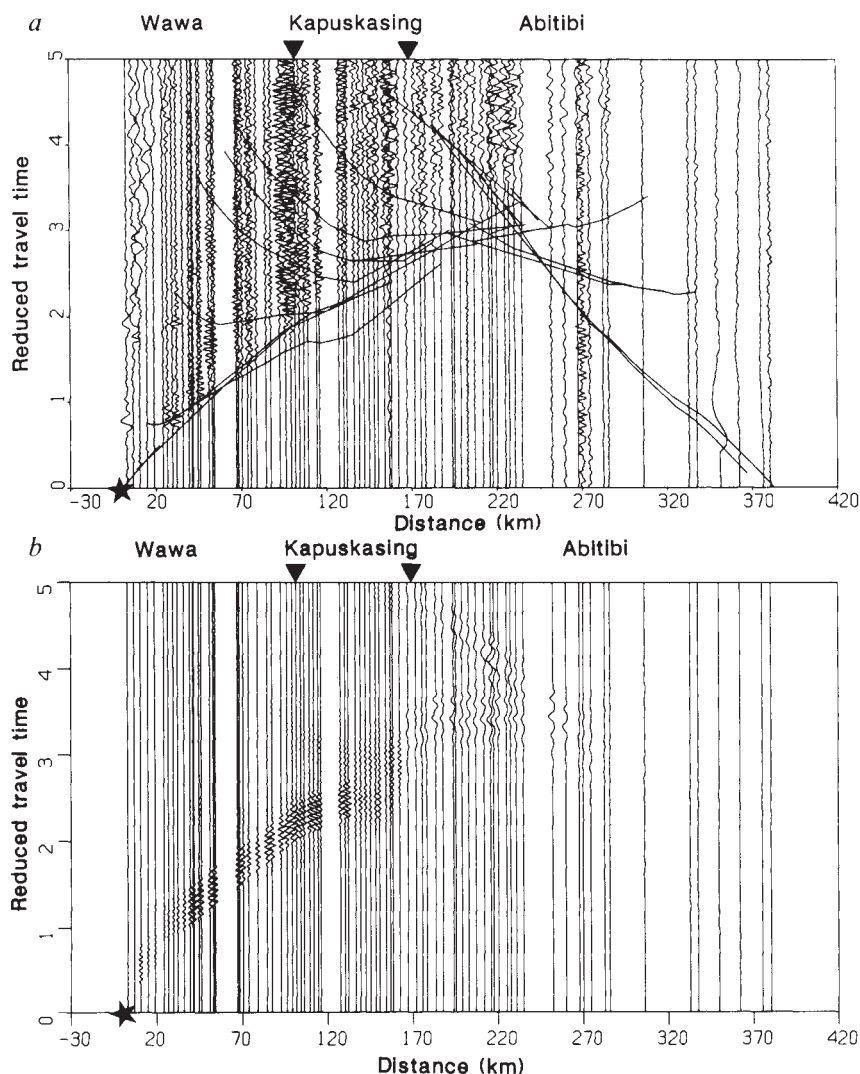
Fig. 4 *a*, Schematic geological cross-section from profile A-A' in Fig. 1 (after Percival and Fountain¹²). Blank areas indicate plutons and orthogneiss, dotted areas indicate granulite facies and V's indicate metavolcanic rocks. *b*, Smoothed isovelocity section of segment of the velocity structure from line 5, resolved on to the cross-section A-A'. I and II are the strong reflectors imaged by the centre shot of line 5.

the true seismograms. One wavelet was chosen for the oscillatory high-frequency upper crustal arrivals and the other for the lower-frequency arrivals from deeper in the crust.

To be able to describe the highly variable nature of the travel-times and amplitudes (Fig. 2*b*), the resultant ray-tracing model was necessarily very complex, so an isovelocity section was constructed by a simple smoothing of a portion of this model (Fig. 4*b*). The near-surface velocities increase progressively from 5.9 km s⁻¹ over the Wawa domal gneiss terrane in the west to 6.5 km s⁻¹ over the southern KSZ, and then decrease abruptly back to 6.25 km s⁻¹ further to the east over the Abitibi greenstone belt (Fig. 4*b*). Material with a velocity of 6.6–6.7 km s⁻¹ and a very low velocity gradient of <0.005 km s⁻¹ per km extends from 20 km depth beneath the Wawa domal gneisses to 4 km under the KSZ. The disturbance of the velocity contours decreases with depth, implying that this velocity anomaly does not reach into the present-day lower crust. The lower-crustal velocities range from 7.2–7.7 km s⁻¹ in the west to 7.2–7.4 km s⁻¹ in the east and the Moho deepens from 48 to 54 km under, and to the west of, the KSZ along this profile. Under the structure, there is much high-frequency intra-crustal scattering, but two highly reflective zones were imaged by the centre shot, at 15–19 km and 23–27 km depth under the KSZ. These show small dips of 6–10° to the west (I and II in Fig. 4*b*). The upper event correlates very well with a prominent event on a preliminary brute-stack of the high-resolution section of the recently acquired Kapuskasing reflection data¹⁶. The lower event, below the penetration depth of the high-resolution data, also shows a highly reflective zone dipping to the west. The Moho is a strong reflector to the east but yields a diffuse response from the deep root in the west.

Our analysis of this seismic data supports the geological model of an intra-cratonic thrust. Other tectonic models have been suggested for the KSZ: a suture¹⁶, a failed arm of a rift junction¹⁷ and a transcurrent fault¹⁸. But the dipping high-velocity anomaly, the thickened crust and the dipping reflectors associated with a downward continuation of the ILCZ provide strong evidence for a thrust model. The isovelocity profile is consistent with the schematic geological cross-section of Percival and Fountain¹¹ (Fig. 4). A high-velocity anomaly is imaged that dips to the west beneath the KSZ. It extends from 20 km depth in

Fig. 3 Comparison of real and synthetic data. *a*, Real data section. Superimposed are the calculated travel-time curves of both the refraction and reflection phases. *b*, Synthetic seismic section obtained by calculating the amplitudes associated with the travel times of the refracted phases of Fig. 2*a* and the reflected phases of Fig. 3*a*. Both sections are plotted with true amplitudes scaled by a distance factor of (offset)^{1.60}. Boundaries of the geological terranes are shown as in Fig. 2.



the west to the surface expression of the KSZ between 65 and 125 km (Fig. 4*b*). This corresponds to the upthrust granulite zone in Fig. 4*a* and could represent the base of a relatively thin Archaean crust uplifted along the Ivanhoe cataclastic zone during an intraplate compressive event. Our results from all five lines indicate that the southern part of the KSZ is the remnant of this event and imply that the thrust fault soles at ~20 km. This agrees well with geobarometric⁶ and preliminary seismic reflection^{19,20} results, and with those from velocity measurements (D. M. Fountain and M. H. Salisbury, personal communication) and gravity modelling⁶. But although this feature is imaged only in the upper half of the present-day crust, the poorer resolution at depth means that a downward extension of the anomaly cannot be ruled out. Alternatively, the ambient pressure, temperature and fluid conditions at such depth may, over time, have altered the rocks so as to eliminate any velocity anomalies. This would project the thrust plane into the lower crust and imply a thicker Archaean crust. Thus, the resolution at depth of these data is not adequate to indicate where the base of the thrust is located. It is expected that joint analysis of the seismic refraction results and newly acquired seismic reflection data will resolve this problem. Over the entire region, the crust thickens from 40–43 km under the greenstone belts to 50–53 km under and to the west of the southern end of the Kapuskasing structure, a characteristic that is also apparent in a time-term analysis of the upper-mantle refraction arrivals²¹. At present, it is unclear whether or not this significant thickening is associated with the trends in the upper and middle-crustal velocities. Thickening of the crust could have been produced by the thrusting of one

section of crust over the other, or it could have developed as an isostatic root to compensate for the presence of anomalous material in the upper crust.

The field program was carried out with funding from Energy, Mines and Resources Canada and the analysis was supported by both that organization and the Natural Sciences and Engineering Research Council of Canada.

Received 20 June; accepted 2 September 1988.

- Percival, J. A. & Card, K. D. *Geology* **11**, 323–326 (1983).
- Percival, J. A., Parish, R. R., Krogh, T. E. & Peterman, Z. E. in *Proc. Kapuskasing Structural Zone Transect Workshop* (ed. West, G. F.) 43–47 (Univ. Toronto, 1988).
- Fountain D. M. & Salisbury, M. H. *Earth planet. Sci. Lett.* **56**, 263–277 (1981).
- Coward, M. P. & Fairhead, J. D. *Tectonophysics* **68**, 31–43 (1980).
- Percival, J. A. & Krogh, T. E. *Can. J. Earth Sci.* **20**, 830–843 (1983).
- Percival, J. A. & McGrath, P. H. *Tectonics* **5**, 553–572 (1986).
- Percival, J. A. in *Reflection Seismology: The Continental Crust* (eds Barazangi, M. & Brown, L.) Am. geophys. Un. Geodyn. Ser. **14**, 135–141 (1986).
- Berry, M. J. & Fuchs, K. *Bull. seismol. Soc. Am.* **63**, 1393–1432 (1973).
- Hall, D. H. & Hajnal, Z. *Bull. seismol. Soc. Am.* **63**, 885–910 (1973).
- Green, A. G., Anderson, N. L. & Stephenson, O. G. *Can. J. Earth Sci.* **16**, 1599–1612 (1979).
- Percival, J. Z. & Fountain, D. M. in *Fluid Movements, Element Transport and the Composition of the Crust* (ed. Bridgwater, D.) (Reidel, Dordrecht, 1988).
- Clowes, R. M. *Geosci. Can.* **11**, 122–126 (1984).
- Zelt, C. A. & Ellis, R. M. *Can. J. Explor. Geophys.* **24**, 16–31 (1988).
- Červený, V., Moltokov, I. A. & Pšenčík, I. *Ray Method in Seismology* (Univ. Karlova, Prague, 1977).
- Ojo, S. B. & Mereu, R. F. *Geophys. J. R. astr. Soc.* **84**, 607–618 (1986).
- Gibb, R. A. & Walcott, R. I. *Earth planet. Sci. Lett.* **10**, 417–422 (1971).
- Burke, K. & Dewey, J. F. *J. Geol.* **81**, 406–433 (1973).
- Watson, J. *Can. J. Earth Sci.* **17**, 866–875 (1983).
- Geis, W. T., Green, A. G. & Cook, F. A. in *Proc. Kapuskasing Structural Zone Transect Workshop* (ed. West, G. F.) 155–165 (Univ. Toronto, 1988).
- Cook, F. A. *Geology* **13**, 368–371 (1985).
- West, G. F. & Northey, D. J. in *Proc. Kapuskasing Structural Zone Transect Workshop* (ed. West, G. F.) 126 (Univ. Toronto, 1988).

Geochemical and climatic effects of increased marine organic carbon burial at the Cenomanian/Turonian boundary

Michael A. Arthur*, Walter E. Dean† & Lisa M. Pratt‡

* Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA

† US Geological Survey, MS 939 Federal Center, Denver, Colorado 80225, USA

‡ Department of Geology, Indiana University, Bloomington, Indiana 47405, USA

Perhaps the most significant event in the Cretaceous record of the carbon isotope composition of carbonate^{1,2}, other than the 1–2.5% negative shift in the carbon isotope composition of calcareous plankton at the Cretaceous/Tertiary boundary³, is the rapid global positive excursion of ~2‰ (¹³C enrichment) which took place between ~91.5 Myr and 90.3 Myr (late Cenomanian to earliest Turonian (C/T boundary event))^{4,5}. This excursion has been attributed to a change in the isotope composition of the marine total dissolved carbon (TDC) reservoir resulting from an increase in rate of burial of ¹³C-depleted organic carbon, which coincided with a major global rise in sea level⁵ during the so-called C/T oceanic anoxic event (OAE)⁶. Here we present new data, from nine localities, which demonstrate that a positive excursion in the carbon isotope composition of organic carbon at or near the C/T boundary^{7,8} is nearly synchronous with that for carbonate and is widespread throughout the Tethys and Atlantic basins (Fig. 1), as well as in more high-latitude epicontinental seas. The postulated increase in the rate of burial of organic carbon may have had a significant effect on CO₂ and O₂ concentrations in the oceans and atmosphere, and consequent effects on global climate and sedimentary facies.

The stratigraphic position of the maximum values of the positive excursion in the carbon isotope composition of organic carbon ($\delta^{13}\text{C}_{\text{org}}$) corresponds closely with the maximum of the excursion in the corresponding quantity for carbonate ($\delta^{13}\text{C}_{\text{carb}}$) (Fig. 2), within the latest Cenomanian and earliest Turonian, over an interval of 0.5–0.8 Myr. Figure 2 shows the excursion within the *Neocardioceras juddi* ammonite zone US-western interior, equivalent to the lower part of the *Whiteinella archeoretacea* planktonic foraminiferal zone and the lower two-thirds of the *Parahabolithus asper* nannofossil zone⁹ and the *Actinocamax plenus* nannofossil subzone of north-western Europe⁴. The $\delta^{13}\text{C}_{\text{org}}$ and $\delta^{13}\text{C}_{\text{carb}}$ excursions both begin sharply in the uppermost *Rotalipora cushmani* planktonic foraminiferal zone (equivalent to the base of the *Sciponoceras gracile* ammonite zone in the US western interior; Fig. 2). The $\delta^{13}\text{C}$ excursion was essentially completed within the time represented by the end of the *Whiteinella archeoretacea* zone. The best age control on the timing of the $\delta^{13}\text{C}_{\text{org}}$ event comes from the US western interior and north-west German sequences^{10,11}, but the $\delta^{13}\text{C}_{\text{carb}}$ excursion has been well dated at a number of other localities⁴.

The marine palaeoenvironments for which C/T $\delta^{13}\text{C}_{\text{org}}$ curves are shown (Fig. 3) include hemipelagic foredelta, pelagic clay-carbonate and pelagic carbonate environments in the US western interior seaway and the European Tethyan realm; and hemipelagic, turbiditic, deep-water, silt-clay upwelling environments in the Atlantic Ocean. In all of these depositional settings, the $\delta^{13}\text{C}_{\text{carb}}$ excursion is 1.5–2‰ and the $\delta^{13}\text{C}_{\text{org}}$ excursion is usually 3.5–4‰. The $\delta^{13}\text{C}_{\text{carb}}$ excursion is generally coincident with the maximum of C_{org} concentrations across the C/T boundary, but C_{org} enrichments occur both somewhat higher and lower than the main $\delta^{13}\text{C}_{\text{carb}}$ peak⁴. The $\delta^{13}\text{C}_{\text{org}}$ excursion likewise does not occur solely in C_{org} -rich intervals, particularly in the US western interior⁸ and north-west German sections, but the $\delta^{13}\text{C}_{\text{org}}$ peak in the Italian C/T sequence does occur entirely

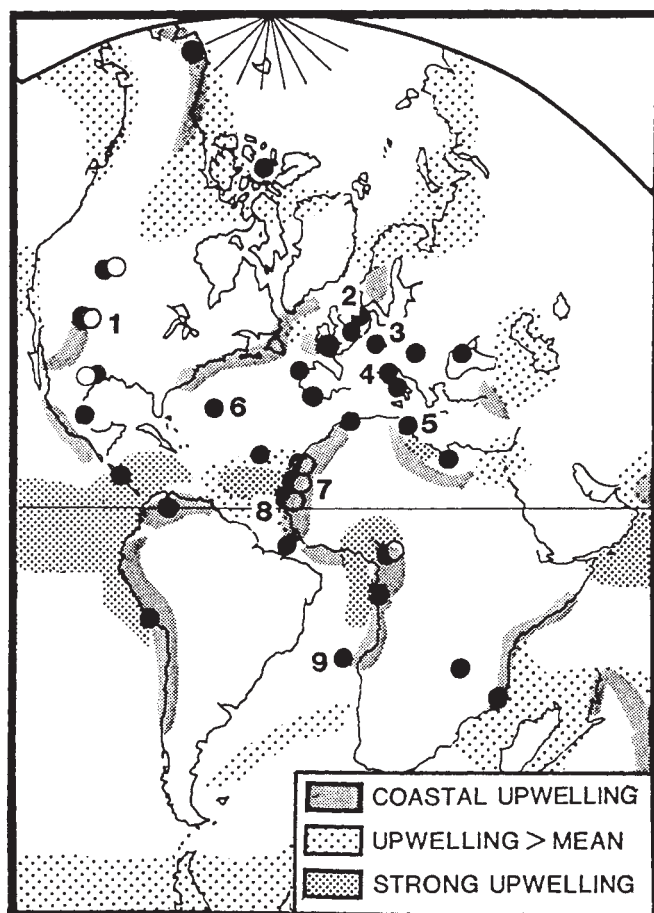


Fig. 1 Map showing locations of Cenomanian/Turonian C_{org} -rich strata (closed circles) and/or laminated, benthic-free strata (open circles). Numbers indicate Cenomanian/Turonian sequences discussed in this paper (Fig. 3). The reconstruction shows that not all C_{org} -rich sequences are characteristic of upwelling zones, but sites such as 7 and 8 within upwelling zones are characterized by greater than normal $\delta^{13}\text{C}_{\text{org}}$ shifts. See ref. 5 for details of construction and source of upwelling information derived from a numerical climate model.

within the so-called Bonarelli Horizon¹², a C_{org} -rich siliceous shale ~1 m thick (Fig. 3).

If the C/T $\delta^{13}\text{C}_{\text{carb}}$ excursion was indeed related to a change in the isotope composition of the oceanic TDC reservoir, the marine $\delta^{13}\text{C}_{\text{org}}$ excursion should be of the same magnitude, assuming that the $\delta^{13}\text{C}_{\text{TDC}}$ is one of the primary controls on $\delta^{13}\text{C}$ of marine C_{org} . But the amplitude of the $\delta^{13}\text{C}_{\text{org}}$ excursion at most localities studied is 3–4‰, and is as much as 6‰ at DSDP sites 367 and 368 (Fig. 3), whereas the maximum amplitude of the $\delta^{13}\text{C}_{\text{carb}}$ excursion at any site is ~2‰ (refs 4, 5). The decay of $\delta^{13}\text{C}_{\text{carb}}$ from relatively more positive values is much more gradual than that apparent for the $\delta^{13}\text{C}_{\text{org}}$ curve (Fig. 2). Thus there may have been other factors, additional to an increase in $\delta^{13}\text{C}_{\text{TDC}}$, responsible for the somewhat different $\delta^{13}\text{C}_{\text{org}}$ and $\delta^{13}\text{C}_{\text{carb}}$ records, such as differential diagenesis or changes in concentration of CO₂ (p_{CO_2}) (refs 13, 14). The widespread C/T $\delta^{13}\text{C}_{\text{org}}$ and $\delta^{13}\text{C}_{\text{carb}}$ excursions clearly are not due to local factors such as diagenesis and changes in source of organic matter. The $\delta^{13}\text{C}_{\text{carb}}$ excursion occurs in pelagic carbonate sequences with or without C_{org} enrichment⁴ but diagenetic overprints of the C_{carb} signal may have occurred in the most C_{org} -rich sequences^{1,10}. For example, bedding-scale fluctuations in $\delta^{13}\text{C}_{\text{carb}}$ superimposed on the excursion are most pronounced in sections with C_{org} -rich marlstones or shales alternating with C_{org} -poor limestone, such as in the Bridge Creek Member of the Greenhorn Limestone (Fig. 2). Neither is differential alter-

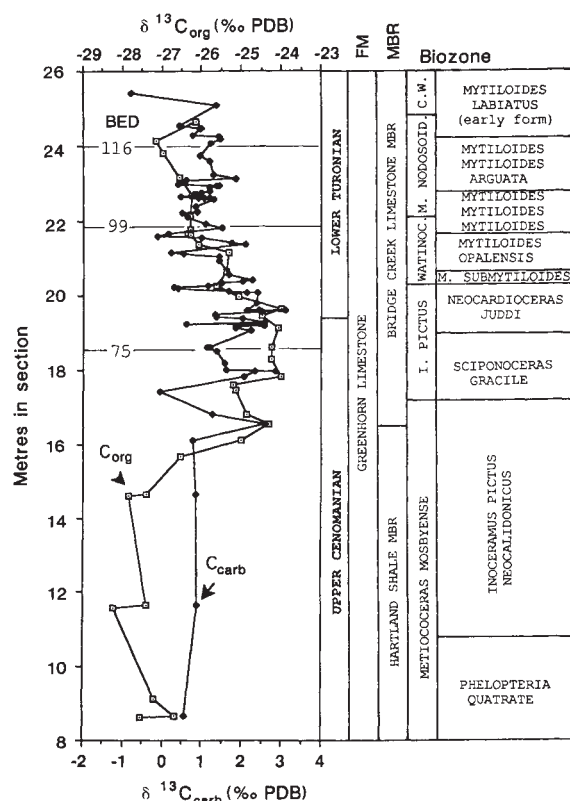


Fig. 2 Plots of $\delta^{13}\text{C}_{\text{org}}$ and $\delta^{13}\text{C}_{\text{carb}}$ versus height of section (in metres) above the base of the Greenhorn Limestone in the Princeton University core, Pueblo, Colorado. Biostratigraphy and age assignments after ref. 10. PDB is the isotope standard. FM denotes formation, MBR denotes member.

ation of the $\delta^{13}\text{C}_{\text{org}}$ signal the probable explanation of the amplitude difference between the curves. None of the sequences illustrated in Fig. 3 have been buried more deeply than 1,200 m, and all are thermally immature with respect to hydrocarbon generation. At most, diagenesis can account for changes of 1–2% in $\delta^{13}\text{C}_{\text{org}}$ (refs 14, 15) but differential alteration of C_{org} in samples just a few centimetres to metres apart in a sequence cannot explain the overall 3–4% shift in $\delta^{13}\text{C}_{\text{org}}$. Similarly, geochemical evidence demonstrates that the background of organic matter throughout the Cenomanian and Turonian at nearly all of the sites examined in this study was predominantly marine and of similar chemical composition^{16,18}.

If the $\delta^{13}\text{C}_{\text{carb}}$ curve is inferred to have reliably recorded the primary surface-water $\delta^{13}\text{C}_{\text{TDC}}$ excursion, there is no obvious explanation for the 1.5–2.5% greater magnitude of the $\delta^{13}\text{C}_{\text{org}}$ than $\delta^{13}\text{C}_{\text{carb}}$ excursion. The peak of the C/T carbon isotopic excursion is the only Cretaceous interval yet sampled in which marine $\delta^{13}\text{C}_{\text{org}}$ values are as heavy as bulk marine organic matter in Miocene and younger marine strata¹⁴. We propose a marked decrease in p_{CO_2} during the C/T OAE as the most likely cause of the amplitude discrepancy between the two types of carbon-isotope curves. Drawing-down of ocean-atmosphere CO_2 levels during the late Cenozoic has been proposed as the principal cause for the shift in marine $\delta^{13}\text{C}_{\text{org}}$ from values of –28‰ in the Cretaceous/Eocene to Miocene/Recent values of –22‰ (refs 13, 14). On the basis of phytoplankton culture experiments and observations of algal photosynthesis in progressively CO_2 -enriched thermal springs it has been argued previously¹⁴ that CO_2 availability (concentration) has an important effect on carbon-isotope fractionation during algal photosynthesis: higher p_{CO_2} leads to an increase in the apparent fractionation factor; that is, marine organic matter becomes relatively more depleted in ^{13}C than does ambient TDC.

Higher burial rates of C_{org} during the C/T OAE would have caused reduction of ocean-atmosphere CO_2 levels, unless the rate of CO_2 release from volcanism and carbonate deposition also proceeded at higher rates. Considering the C/T sea-level highstand and extensive areas of epicontinental seas at this time, rates of continental weathering were probably somewhat reduced and rates of carbonate deposition somewhat elevated in the late Cenomanian^{2,19}. On the other hand, the stratigraphic distribution of volcanic ash beds and the ages of plutonic rocks indicate that mid-plate and explosive plate-margin volcanism, and consequent CO_2 outgassing, was greatest during the Aptian/Albian (100–120 Myr) and that the C/T sea-level highstand accompanied a minimum of volcanic activity²⁰. We conclude, therefore, that significant drawdown of CO_2 occurred during the C/T event. This short-term reduction of CO_2 availability as the result of higher productivity and/or higher C_{org} burial rate is proposed as the principal cause for the larger positive shift in $\delta^{13}\text{C}_{\text{org}}$ than that in pelagic carbonate. That such positive $\delta^{13}\text{C}_{\text{org}}$ excursions do not accompany all C_{org} burial events in the geological record is illustrated by the Toarcian (early Jurassic) episode during which a positive excursion in $\delta^{13}\text{C}_{\text{carb}}$ is accompanied by a pronounced negative excursion in $\delta^{13}\text{C}_{\text{org}}$ (ref. 21).

In the highly C_{org} -rich strata (up to 50% C_{org}) deposited off northwest Africa (DSDP sites 367 and 368) the $\delta^{13}\text{C}_{\text{org}}$ excursion is 5.5–6‰ (Fig. 3). We attribute this 2% greater magnitude of the excursion at sites 367 and 368, compared to other studied localities, to a water-column signal unique to upwelling shelves and other high-productivity settings during the C/T episode. Apparently, upwelling of nutrients and high primary productivity resulted in further CO_2 depletion and even less carbon-isotope fractionation by phytoplankton.

The timing and nature of the positive isotope excursions in both C_{carb} and C_{org} have profound implications for rates of C_{org} burial during the C/T event and consequent potential effects on p_{CO_2} and p_{O_2} in the atmosphere. Our preliminary model calculations using the carbon isotope constraints provided by our data, and following the usual isotope mass-balance modelling procedures²², suggest that the global rate of C_{org} burial increased by a factor of 1.3 during the main C/T episode over that prevailing during preceding Cenomanian time. The inferred increase in C_{org} burial rates is supported by a variety of sedimentological, palaeontological and geochemical evidence^{4–7,16,23,24}. The peak C/T C_{org} burial episode is estimated to have lasted at least 0.5 Myr (refs 4, 5). If we assume that the riverine flux of dissolved carbon to the oceans throughout the C/T interval was the same as at present (3.5×10^{13} mol C per year (ref. 19)), the calculated excess burial of C_{org} above the Cenomanian background was about 0.32×10^{13} mol C per year during the C/T event, or a total excess of about 1.6×10^{18} mol C for the 0.5-Myr episode.

At that rate of excess- C_{org} burial, it would take only ~10,000–20,000 yr to consume all of the CO_2 in the present atmosphere, assuming that rates of CO_2 production and consumption remained constant. But it has been suggested that the middle-to-late Cretaceous atmosphere was characterized by a p_{CO_2} that was 4–18 times higher than at present^{2,19,25}. If we assume that p_{CO_2} was eight times higher than at present, the increased rate of C_{org} burial during the C/T event was still sufficient to strip the atmosphere of CO_2 within ~120,000 yr.

Although these simple calculations ignore the influence of the ocean TDC reservoir, they suggest that in all likelihood there was a significant reduction in p_{CO_2} as a result of the C/T C_{org} burial event. This has two major implications: (1) that the decreased availability of dissolved CO_2 affected the isotope fractionation of C_{org} by phytoplankton as argued above, and (2) that the decrease in p_{CO_2} probably induced significant global climatic cooling during and after the C/T event. The latter implication is a prediction that is testable but, at present, there are few supporting oxygen isotope data. However, oxygen

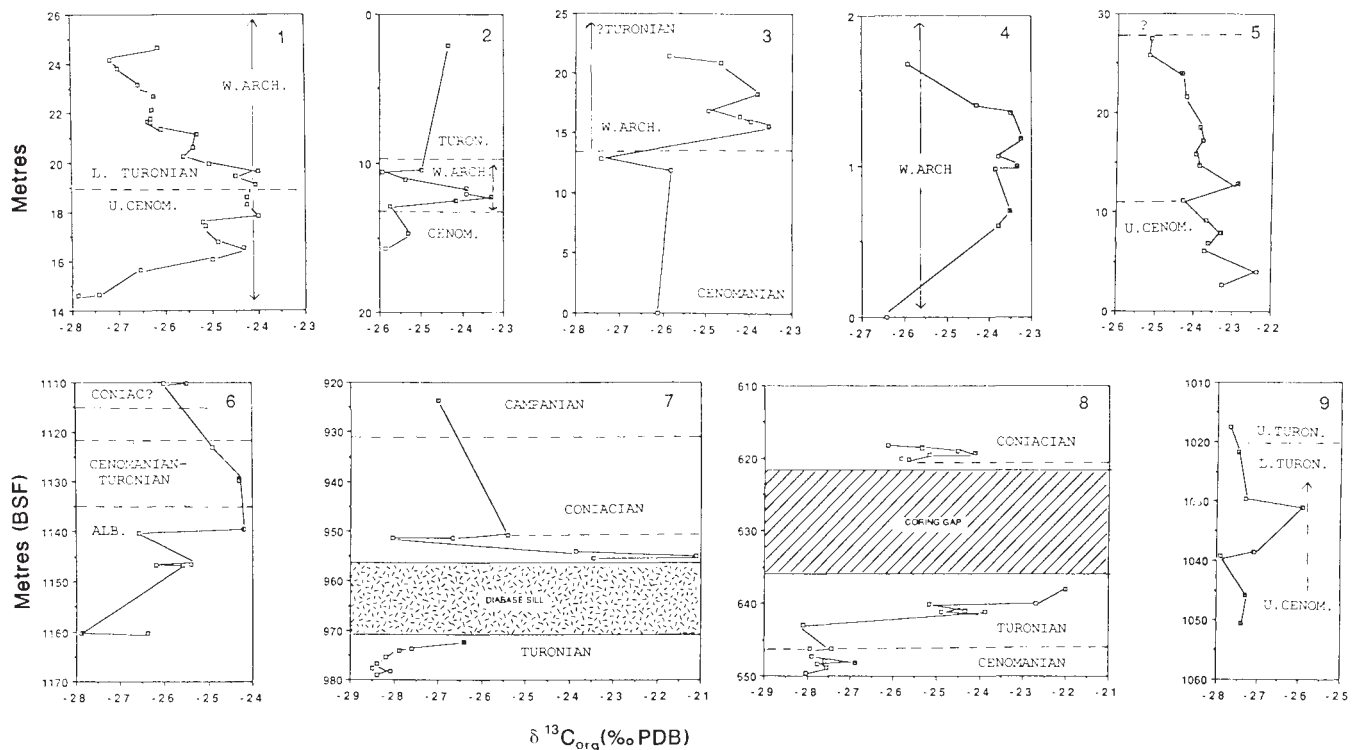


Fig. 3 Plots of $\delta^{13}\text{C}_{\text{org}}$ versus depth (in metres below sea floor (BSF)) for Cenomanian/Turonian sections for the numbered sequences shown in Fig. 1. Extent of *Whiteinella archeoetacea* planktonic foraminiferal zone is indicated where known. Background information on biostratigraphy, lithology, and geochemistry for each of the sequences can be found in the relevant DSDP Initial Reports Volumes and/or in the following selected references: (1) Bridge Creek Limestone, Pueblo, Colorado: refs 7, 8, 10, 18; (2) Danish Central Graben; ref. 6; (3) Wunstorf, north Germany: ref. 6; (4) North-central Italy (Bonarelli Horizon): refs 12, 17; (5) Bahloul Formation, Tunisia: ref. 6; (6) DSDP site 603; (7) DSDP site 368; (8) DSDP site 367; (9) DSDP site 530.

isotope compositions of inoceramids in chalk sequences from the UK and north-west Germany² suggest that there was an increase in $\delta^{18}\text{O}$ of 2–3 ‰ during the Turonian, indicating cooling of 8–13 °C there. In addition, micropalaeontological data suggest that there was an incursion of boreal elements into lower latitudes, such as offshore of northwest Africa, during the Turonian²⁶. The drop in p_{CO_2} could have provided positive feedback to induce climatic cooling at high latitudes and stimulate upwelling of colder deep waters at low latitudes. Higher biotic turnover rates²⁷ would be an obvious result of such chemical and climatic changes, thereby providing one explanation for the C/T extinction events²⁸.

An additional consideration is the potential effect that enhanced C_{org} burial at the C/T boundary had on the atmospheric oxygen reservoir. We estimate that increased C_{org} burial of 1.6×10^{18} mol C over 0.5 Myr across the C/T boundary would have resulted in an equal addition of O_2 to the oxygen reservoir, assuming a CO_2 -to- O_2 photosynthetic ratio of 1:1. This amount is more than 40% of the present mass of atmospheric O_2 and represents a significant p_{O_2} increase over a brief time interval. In addition, the p_{O_2} excess apparently was not offset by a change in the sulphur cycle, as occurs during some other periods of more rapid C_{org} burial²⁸. The $\delta^{34}\text{S}$ of sea water was increasing at a rate of about 0.8 ‰ Myr^{-1} during the Cenomanian and early Turonian², indicating that there was a net transfer of sulphur to the sulphide reservoir and a net release of O_2 .

An increase in p_{O_2} accompanied by cooling following the late stage of the C/T event would have increased the oxidizing capacity of oceanic deep waters by 50% or more, assuming that the rate of O_2 consumption during weathering did not change. This is perhaps part of the explanation for the widespread onset of multicoloured, oxidized pelagic clay deposition²⁹ on the deeper Atlantic sea floor and for the switch to oxidized, reddish pelagic carbonates in the Tethys during the Turonian¹², follow-

ing prolonged deposition of dark-coloured, C_{org} -rich lithologies in both regions.

Enhanced C_{org} burial during the Cenomanian and Turonian had far-reaching consequences. The biological and geochemical evidence for higher C_{org} burial is overwhelming. Evidence for decreased p_{CO_2} in the atmosphere and oceans, and attendant global climatic cooling, also is available. Some evidence for increased p_{O_2} in the atmosphere³⁰ and the oceans exists but needs additional documentation. Clearly, higher-resolution records of carbon, oxygen and sulphur stable-isotope compositions are needed to further test and define these relationships. Such events also illustrate mechanisms by which changes in rates of tectonism, outgassing and consequent sea-level changes can induce significant, rapid palaeoenvironmental changes and biotic turnover.

Received 11 April; accepted 16 August 1988.

- Scholle, P. A. & Arthur, M. A. *Bull. Am. Ass. Petrol. Geol.* **64**, 67–87 (1980).
- Arthur, M. A., Dean, W. E. & Schlanger, S. O. in *The Carbon Cycle and Atmospheric CO_2 : Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 504–530 (Am. Geophys. Un. Washington, Monograph 32, 1985).
- Zachos, J. C. & Arthur, M. A. *Paleoceanography* **1**, 5–26 (1986).
- Schlanger, S. O., Arthur, M. A., Jenkyns, H. C. & Scholle, P. A. in *Marine Petroleum Source Rocks* (eds Brooks, J. & Fleet, A. J.) 371–399 (Geol. Soc. London, spec. Publ. 26, 1987).
- Arthur, M. A., Schlanger, S. O. & Jenkyns, H. C. in *Marine Petroleum Source Rocks* (eds Brooks, J. & Fleet, A. J.) 401–420 (Geol. Soc. London, spec. Publ. 26, 1987).
- Schlanger, S. O. & Jenkyns, H. C. *Geol. Mijnbouw* **55**, 179–184 (1976).
- Pratt, L. M. in *Fine-Grained Deposits and Biofacies of the Cretaceous Western Interior Seaway: Evidence of Cyclic Sedimentary Processes* (eds Pratt, L. M., Kauffman, E. G. & Zelt, F. B.) 38–48 (Soc. Econ. Paleontol. Miner., Tulsa, 1985).
- Pratt, L. M. & Threlkeld, C. N. in *Mesozoic of Middle North America* (eds Stott, D. F. & Glass, D. J.) 305–312 (Can. Soc. Petrol. Geol., Memoir 9, 1984).
- Bralower, T. J., *Paleoceanography* (in the press).
- Fine-Grained Deposits and Biofacies of the Cretaceous Western Interior Seaway: Evidence of Cyclic Sedimentary Processes* (eds Pratt, L. M., Kauffman, E. G. & Zelt, F. B.) (Soc. Econ. Paleontol. Miner., Tulsa, 1985).
- Ernst, G., Schmid, F. & Klischies, G. in *Aspekte de Kreide Europas* (ed. Wiedmann, J.) 11–46 (Int. Un. geol. Sci. Ser. A, No. 6, 1979).
- Arthur, M. A. & Premoli-Silva, I. in *Nature and Origin of Cretaceous Carbon-Rich Facies* (eds Schlanger, S. O. & Cita, M. B.) 7–54 (Academic, London, 1982).
- Arthur, M. A., Dean, W. E. & Claypool, G. E. *Nature* **315**, 216–218 (1985).
- Dean, W. E., Arthur, M. A. & Claypool, G. E. *Mar. Geol.* **70**, 119–157 (1986).

15. Deines, P. in *Handbook of Environmental Isotope Geochemistry*, Vol. 1, (eds Fritz, P. & Fontes, J. C.) 329–406 (Elsevier, Amsterdam, 1980).
16. Herbin, J. P. *et al.*, in *North Atlantic Palaeoceanography* (eds Summerhayes, C. P. & Shackleton, N. J.) 389–422 (Geol. Soc. Lond. spec. Publ. 21, 1986).
17. Graas, G. van, Viets, T. C., de Leeuw, J. W. & Schenck, P. A. *Geochim. cosmochim. Acta* **47**, 1051–1059.
18. Pratt, L. M. *Bull. Am. Ass. Petrol. Geol.* **68**, 1146–1159 (1984).
19. Lasaga, A. C., Berner, R. A. & Garrels, R. M. in *Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 397–411 (Am. Geophys. Un. Washington, Monograph 32, 1985).
20. Schlanger, S. O. in *Mesozoic and Cenozoic Oceans* (ed. Hsü, K. J.) (Am. Geophys. Un. Geodyn. Ser. 15, 1986).
21. Jenkyns, H. C. & Clayton, C. J. *Sedimentology* **33**, 87–106 (1986).
22. Berner, R. A. *Am. J. Sci.* **287**, 177–196 (1987).
23. Jenkyns, H. C. *J. geol. Soc. Lond.* **137**, 171–188 (1980).
24. De Graciansky, P. C. *et al.* in *Marine Petroleum Source Rocks* (eds Brooks, J. & Fleet, A. J.) 317–344 (Geol. Soc. Lond., spec. Publ. 26, 1987).
25. Barron, E. J. & Washington, W. M. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 504–529 (Am. Geophys. Un. Washington, Monograph 32, 1985).
26. Kuhnt, W., Thürow, J. & Wiedman, J. *Mitt. Geol.-Paläontol. Inst. Univ. Hamburg, SCOPE/UNEP* **60**, 205–246 (1986).
27. Hut, P. *et al. Nature* **329**, 118–126 (1987).
28. Kump, L. R. & Garrels, R. M. *Am. J. Sci.* **286**, 337–360 (1986).
29. Jansa, L. F., Enos, P., Tucholke, B. E., Gradstein, F. M. & Sheridan, R. E. in *Deep Drilling Results in the Atlantic Ocean: Continental Margins and Palaeoenvironments* (eds Talwani, M., Hay, W. W. & Ryan, W. B. F.) 1–5 (Am. Geophys. Un. Maurice Ewing Ser. 3, 1979).
30. Berner, R. A. & Landis, G. P. *Geol. Soc. Am. Abs. Prog.* **19**, 587 (1987).

Extracting environmental information from planktonic foraminiferal $\delta^{13}\text{C}$ data

Howard J. Spero & Douglas F. Williams

Ocean Sciences Division, Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

The oxygen isotopic composition of foraminiferal shell calcite varies as a function of temperature and the $^{18}\text{O}/^{16}\text{O}$ ratio of seawater during calcification. This relationship has provided the basis for numerous studies of past climate changes, continental glaciation and changes in oceanic current systems¹. In contrast, the variables affecting the $\delta^{13}\text{C}$ composition of foraminiferal calcite are poorly understood. It has been proposed that the $\delta^{13}\text{C}$ value of planktonic foraminifera is a function of the $^{13}\text{C}/^{12}\text{C}$ ratio of sea water ΣCO_2 and of 'vital effects' related to metabolic processes such as respiration and symbiont photosynthesis^{1,2}. In this paper we demonstrate that the $\delta^{13}\text{C}$ value of *Orbulina universa* shell calcite is a function of symbiont photosynthetic activity and irradiance levels based on laboratory experiments with living specimens. Environmental temperature has no significant effect on the $\delta^{13}\text{C}$ value, whereas the $\delta^{18}\text{O}$ value of the shell varies as a function of temperature but is independent of symbiont photosynthesis. The ability to relate $\delta^{13}\text{C}$ data with the environmental parameter of light may provide new insights into dynamic processes occurring in palaeoceanic euphotic zones.

Orbulina universa secretes a large terminal, spherical shell which comprises 90–95% of its calcite³. Numerous dinoflagellate symbionts are located on spines in a halo around the shell^{4,5} where maximum photosynthetic rates (P_{max}) occur at light intensities greater than $386 \mu\text{Einst m}^{-2} \text{s}^{-1}$ (ref. 4). For the laboratory experiments, specimens were maintained under four light regimes between dark and P_{max} intensities to quantify the effect of the full range of symbiont photosynthetic rates on the $\delta^{13}\text{C}$ value of the shell. A control group maintained at near P_{max} intensities was treated with the photosynthetic inhibitor 3(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) to eliminate the symbiotic algae⁶ and examine the effect of light on the isotopic composition of aposymbiotic foraminifera. Previous research demonstrated that DCMU does not affect the ability of the foraminifer to secrete chambers or undergo gametogenesis^{5,6}. Culture experiments were repeated at temperatures 16.5 ± 1.0 , 20 ± 0.2 and 25 ± 0.2 °C, to represent the natural temperature range of *O. universa*⁷. Sea water used for culturing was changed

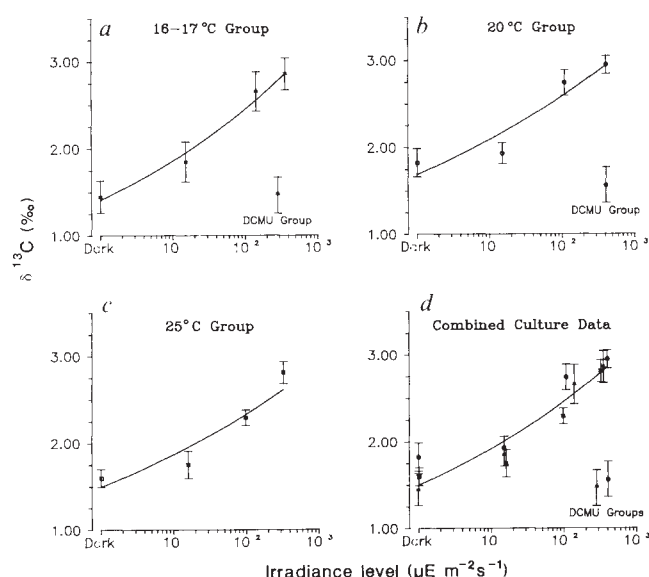


Fig. 1 *a–c*, $\delta^{13}\text{C}$ versus light-intensity relationships for cultured *O. universa*. The values for $\delta^{13}\text{C}$ increase with irradiance up to intensities of $\sim 386 \mu\text{Einst m}^{-2} \text{s}^{-1}$, the symbiont photosynthetic P_{max} (ref. 8). DCMU-group data are from aposymbiotic *O. universa*. Regression lines, using a power regression function, give correlation coefficients (r^2) of 0.89, 0.81 and 0.83 for temperature groups 16.5, 20 and 25 °C, respectively. These regressions are not significantly different. *d*, The composite $\delta^{13}\text{C}$ versus light relationship for *O. universa* from all temperature groups as described by the regression equation given in the text.

Table 1 Mean (± 1 s.d.) stable isotope data (‰) for ambient surface sea water and 48-h-old culture water analysed using established techniques^{20,21}

Group	Temperature (°C)	$\delta^{18}\text{O}_{\text{SMOW}}$ (‰)	$\delta^{13}\text{C}_{\text{PDB}}$ (‰)
(A) Ambient sea water	16.5–21.5	-0.53 ± 0.06 (6)	$+1.67 \pm 0.08$ (5)
(B)	16.5	-0.50 ± 0.01 (2)	$+1.44 \pm 0.07$ (2)
48-h culture water	20	NA	$+1.40 \pm 0.16$ (4)
	25	-0.42 ± 0.07 (5)	$+1.45 \pm 0.23$ (6)
$\bar{X}$ culture water (A and B) combined:		-0.48 ± 0.07	$+1.50 \pm 0.19$
(C) $\Delta A - B$		0.09	0.27

O. universa was maintained in 100 ml of 0.45- μm filtered sea water. The water in each culture was changed every 48 h and randomly sampled for $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ isotopic analysis. $\delta^{18}\text{O}$ values are relative to standard mean ocean water (SMOW); $\delta^{13}\text{C}$ values relative to PDB (radiocarbon dating standard). The data on the culture water combine isotopic analyses of water from all light groups. The numbers in parentheses are the numbers of samples analysed; NA—no samples analysed. The isotopic difference between ambient surface and culture sea water ($\Delta A - B$) demonstrate that the isotopic composition of the culture water is not significantly affected by the culturing procedures used.

every 48 h. Isotopic analyses of randomly selected culture water samples indicate that 48-h old water is isotopically similar to ambient sea water (Table 1).

The $\delta^{13}\text{C}$ values of cultured *O. universa* are highly correlated with irradiance at ambient temperatures between 16.5 and 25 °C (Table 2, Fig. 1) due to changes in the photosynthetic rate of the symbiotic algae⁸. Under maximum light intensities, specimens are enriched in ^{13}C relative to lower light intensities. In the dark, specimens are depleted in ^{13}C by ~ 1.20 – 1.50 ‰ relative to the highest light-intensity group. The effect of symbiont photosynthetic activity is clearly shown by the group treated with DCMU. Aposymbiotic *O. universa* grown at near- P_{max} light intensities have $\delta^{13}\text{C}$ values identical to symbiotic *O. universa*

maintained in the dark (Fig. 1a, b). These results support the hypothesis that the $^{13}\text{C}/^{12}\text{C}$ ratio of the spherical chamber is controlled by the photosynthetic activity of the symbiotic algae⁸ rather than some direct relationship between light and carbon isotope incorporation. Furthermore, environmental temperature has little effect on the $\delta^{13}\text{C}$ value of *O. universa* because the $\delta^{13}\text{C}$ -irradiance relationships are similar for the three temperature groups (Fig. 1a-c).

Several investigators have reported *O. universa* with larger diameter spheres from sediment samples were enriched in ^{13}C relative to smaller shells⁹⁻¹¹. The size data presented in Table 2 demonstrate that the observed $\delta^{13}\text{C}$ -irradiance relationship is not a function of shell diameter. Laboratory data have demonstrated, however, that the mean diameter of *O. universa* increases with culture illumination, a function of increased symbiont photosynthetic rate⁸. Therefore, field observations of enriched $\delta^{13}\text{C}$ values in large *O. universa* shells may reflect calcification at high irradiance levels in the oceanic euphotic zone.

The $\delta^{18}\text{O}$ values of *O. universa* (Table 2) demonstrate that symbiont photosynthetic activity does not significantly affect the $^{18}\text{O}/^{16}\text{O}$ ratio of the shell in the temperature range 16.5–25 °C. Calculations of the predicted temperature based on equilibrium precipitation¹² give mean temperatures of 18.2 ± 1.5 , 20.7 ± 1.8 and 25.2 ± 0.9 °C for the 16.5, 20 and 25 °C groups, respectively. These data support previous laboratory studies indicating that *O. universa* secretes calcite in oxygen equilibrium^{13,14}.

These results indicate that the $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values of *O. universa* represent mutually exclusive signals which are controlled by the environmental parameters, light and temperature. By combining these data, it may be possible to relate light levels with temperature in the euphotic zone. The relationship between $\delta^{13}\text{C}$ and irradiance (I , $\mu\text{Einst m}^{-2} \text{ s}^{-1}$) (Fig. 1d) can be described by the power regression

$$\delta^{13}\text{C} = 1.50 \times I^{0.107} \quad (r^2 = 0.87)$$

for irradiance levels between dark and the dinoflagellate symbiont P_{max} intensity, $386 \mu\text{Einst m}^{-2} \text{ s}^{-1}$. At brighter intensities, the $\delta^{13}\text{C}$ value should not enrich further⁸ because the symbiont photosynthetic rate is maximum. Because the $\delta^{18}\text{O}$ value of the shell is a function of temperature, if the $^{18}\text{O}/^{16}\text{O}$ ratio of the water can be measured or estimated, it should be possible to relate light levels to temperature in the thermocline and mixed layer by combining the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ isotope values of *O. universa*.

From these relationships, we postulate that *O. universa* calcifying in the more illuminated mixed layer of the ocean would have enriched $\delta^{13}\text{C}$ values relative to individuals from low light environments of the thermocline. Similarly, individuals in the upper mixed layer would have depleted $\delta^{18}\text{O}$ values (corresponding to warmer temperatures) relative to individuals calcifying deep in the seasonal thermocline. Furthermore, in a sediment assemblage of fossil shells, the range of $\delta^{18}\text{O}$ values at maximum and minimum $\delta^{13}\text{C}$ values would correspond to the seasonal range of temperatures in the mixed layer and deep seasonal thermocline respectively.

The mean $\delta^{13}\text{C}$ value of all dark and DCMU culture groups combined, $+1.55 \pm 0.20\%$, is virtually identical to the mean culture water ΣCO_2 , $+1.50 \pm 0.19\%$. In these experimental groups with no symbiont photosynthetic activity, CO_2 for calcification must come from either the respiratory or seawater ΣCO_2 pool. If respired CO_2 is the primary source, the shells should have more negative $\delta^{13}\text{C}$ values because organic matter is very depleted in ^{13}C ($\approx -20\%$)¹⁵. We suggest, however, that diffusional processes around the calcifying shell are rapid enough to dilute the respiratory CO_2 pool and result in a shell $^{13}\text{C}/^{12}\text{C}$ ratio similar to that of ΣCO_2 (ref. 16). Therefore, we propose that the most depleted $\delta^{13}\text{C}$ values in a natural sediment assemblage of *O. universa* correspond to calcification under very low light conditions, and provide a measure of the $\delta^{13}\text{C}$ of ΣCO_2 at the base of the euphotic zone.

Table 2 Stable isotope and mean diameter (± 1 s.d.) data for laboratory-grown *Orbulina universa*

Irradiance level ($\mu\text{Einst m}^{-2} \text{ s}^{-1}$)	N	Mean diameter (μm)	$\delta^{13}\text{C}$ (‰, PDB)	$\delta^{18}\text{O}$ (‰, PDB)
16.5 $\pm$ 1 °C Group				
357	4	464 $\pm$ 68	2.87	-1.04
	3	522 $\pm$ 56	3.00	-0.80
	5	486 $\pm$ 52	2.91	-0.77
	4	519 $\pm$ 48	2.54	-1.19
	5	470 $\pm$ 35	2.97	-1.28
141	5	481 $\pm$ 25	2.40	-1.05
	6	464 $\pm$ 40	2.74	-1.56
15	2	471 $\pm$ 31	2.83	-0.57
	5	455 $\pm$ 56	1.76	-0.71
	4	519 $\pm$ 43	1.68	-1.00
Dark	5	483 $\pm$ 66	2.11	-0.62
	3	443 $\pm$ 19	1.19	-1.43
	3	459 $\pm$ 7	1.60	-1.54
DCMU (282)	3	497 $\pm$ 20	1.56	-1.04
	3	471 $\pm$ 31	1.44	-1.92
	3	471 $\pm$ 49	1.31	-1.48
DCMU (282)	2	524 $\pm$ 31	1.70	-1.36
	3	490 $\pm$ 35	1.38	-1.49
20 $\pm$ 0.2 °C Group				
407	2	519 $\pm$ 13	3.03	-1.85
	2	541 $\pm$ 44	2.98	-2.28
	2	502 $\pm$ 12	2.83	-2.05
109	2	550 $\pm$ 6	2.70	-1.94
	2	583 $\pm$ 47	2.83	-1.75
	3	461 $\pm$ 20	2.55	-2.28
15	1	572	2.88	-1.70
	2	537 $\pm$ 25	2.04	-1.01
	4	430 $\pm$ 37	1.95	-1.21
Dark	2	581 $\pm$ 49	1.80	-2.29
	3	499 $\pm$ 10	1.71	-1.74
	4	438 $\pm$ 27	1.94	-1.33
DCMU (407)	3	478 $\pm$ 21	1.42	-2.18
	3	478 $\pm$ 27	1.71	-1.13
25 $\pm$ 0.2 °C Group				
329	2	493 $\pm$ 13	2.95	-2.73
	2	436 $\pm$ 19	2.78	-2.95
	2	515 $\pm$ 6	2.70	-2.73
99	2	524 $\pm$ 19	2.18	-3.19
	2	515 $\pm$ 18	2.35	-2.73
	2	498 $\pm$ 6	2.26	-3.53
16	2	542 $\pm$ 6	2.38	-3.14
	2	546 $\pm$ 62	1.72	-2.99
	3	487 $\pm$ 19	1.92	-2.70
Dark	3	449 $\pm$ 18	1.61	-2.98
	3	461 $\pm$ 22	1.48	-2.66
	4	420 $\pm$ 52	1.68	-2.87
Dark	3	503 $\pm$ 16	1.63	-2.99

Trochospiral-stage *O. universa* were hand collected by scuba divers in the San Pedro Basin, 2 km NNE of the Catalina Marine Science Center, Santa Catalina Island, California. Specimens were cultured in the laboratory where the spherical chamber was secreted under controlled light and temperature conditions. Each foraminifer was fed a one-day-old *Artemia* nauplius every other day. Culture illumination was provided by GE F24 T12 CW-HO fluorescent bulbs. About five days after sphere formation, the foraminifer releases gametes into the water resulting in a shell devoid of cytoplasm^{3,22}. Empty post-gametogenic shells were collected, cracked open and the juvenile trochospiral shells were removed. Shells (1–5) (50–110 μg) were used for each isotopic analysis following standard techniques²³. Note that specimens in the DCMU groups are aposymbiotic. *N* is the number of shells analysed for each measurement.

To extract environmental information from fossil foraminiferal assemblages using these relationships, it is necessary to analyse individual *O. universa* because each specimen records specific water-column information which is averaged in multi-specimen analyses. Only two such isotope

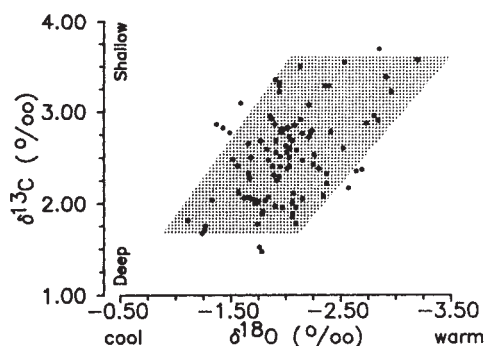


Fig. 2 Published $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ data from analyses of individual shells of *O. universa* from the western equatorial Pacific Ocean^{17,18}. These data display the predicted $\delta^{13}\text{C}$ - $\delta^{18}\text{O}$ relationship for a fossil *O. universa* assemblage, that is, the most enriched $\delta^{13}\text{C}$ values (high irradiance level) correspond with the more depleted range of $\delta^{18}\text{O}$ values (warm water) and vice versa. The stippled region delineates this isotopic trend. The absolute $\delta^{13}\text{C}$ range of these data are in good agreement with the range of our laboratory data, 1.31–3.03‰.

data sets are available for *O. universa*, both from a single western equatorial Pacific box core^{17,18}. These data show that the predicted $\delta^{13}\text{C}$ - $\delta^{18}\text{O}$ trend does exist—enriched $\delta^{13}\text{C}$ values correspond with a more depleted range of $\delta^{18}\text{O}$ values than the values corresponding with depleted $\delta^{13}\text{C}$ values (Fig. 2). The absolute $\delta^{13}\text{C}$ range for *O. universa* in this box core (1.48–3.60‰) is in good agreement with the range of our laboratory data (1.31–3.02‰). Furthermore, the mean $\delta^{13}\text{C}$ value of mixed layer ΣCO_2 at stations near this core site (Stations 246 and 251)¹⁹ is 1.42‰, close to the minimum $\delta^{13}\text{C}$ values for *Orbulina*.

It is premature to extrapolate the irradiance- $\delta^{13}\text{C}$ relationship in *O. universa* to other symbiotic species such as *Globigerinoides sacculifer* or *G. ruber*. Additional laboratory experiments are needed to determine whether these species also display a quantifiable relationship. Nevertheless, the ability to extract environmental information from both carbon and oxygen isotopes in *O. universa* may enable palaeoceanographers to make new inferences about upper ocean processes.

We thank W. Berger and S. Burke for providing us with the original data set of Killingley *et al.*¹⁷, N. Healy-Williams and R. Thunell for comments, and C. Cline, W. Barrett and the staff of the Catalina Marine Science Center for field assistance. This study was supported by the Department of Geological Sciences and the Stable Isotope Laboratory, University of South Carolina.

Received 30 May; accepted 30 August 1988.

- Duplessy, J.-C. in *Climatic Change* (ed. Gribbin, J.) 46–67 (Cambridge Univ. Press, Cambridge, 1978).
- Erez, J. & Honjo, S. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 129–156 (1981).
- Spero, H. J. *Mar. Biol.* **99**, 9–20 (1988).
- Spero, H. J. & Parker, S. L. *J. foram. Res.* **15**, 273–281 (1985).
- Spero, H. J. *J. Phycol.* **23**, 307–317 (1987).
- Bé, A. W. H., Spero, H. J. & Anderson, O. R. *Mar. Biol.* **70**, 73–86 (1982).
- Bé, A. W. H., Harrison, S. M. & Lott, L. *Micropalaeontology* **19**, 150–192 (1973).
- Spero, H. J. & DeNiro, M. J. *Symbiosis* **4**, 213–228 (1987).
- Weiner, S. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **17**, 149–156 (1975).
- Berger, W. H., Killingley, J. S. & Vincent, E. *Oceanologica Acta* **1**, 203–316 (1978).
- Williams, D. F., Bé, A. W. H. & Fairbanks, R. G. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 71–201 (1981).
- Epstein, S., Buchsbaum, R., Lowenstam, H. A. & Urey, H. C. *Bull. geol. Soc. Am.* **64**, 1315–1326 (1953).
- Bouvier-Soumagnac, Y. & Duplessy, J.-C. *J. foram. Res.* **15**, 302–320 (1985).
- Bouvier-Soumagnac, Y., Duplessy, J.-C. & Bé, A. W. H. *Oceanologica Acta* **9**, 519–522 (1986).
- DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **42**, 495–506 (1978).
- Jørgensen, B. B., Erez, J., Revsbech, N. P. & Cohen, Y. *Limnol. Oceanogr.* **30**, 1253–1267 (1985).
- Killingley, J. S., Johnson, R. F. & Berger, W. H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 193–204 (1981).
- Schiffelbein, P. & Hills, S. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **48**, 197–213 (1984).
- Kroonnick, P. *Deep Sea Res.* **32**, 57–84 (1985).
- Fairbanks, R. G., Wiebe, P. H. & Bé, A. W. H. *Science* **207**, 61–63 (1980).
- Grossman, E. L. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **47**, 301–327 (1984).
- Caron, D. A., Faber, W. W. & Bé, A. W. H. *J. mar. Biol. Ass. U.K.* **7**, 343–358 (1987).
- Williams, D. F., Sommer, M. A., II & Bender, M. L. *Earth planet. Sci. Lett.* **36**, 391–403 (1977).

Traditionality of mating-site preferences in a coral reef fish

Robert R. Warner

Department of Biological Sciences and Marine Science Institute, University of California, Santa Barbara, California 93106, USA

Mating systems are often viewed as a result of females choosing where, when and with whom to mate^{1,2}, and of male responses to female behaviour^{3,4}. If the distribution of females reflects their response to vital resources for survival and reproduction, then a careful study of these resources should yield predictions about the emergent mating system. Males and females could, however, simply gather at traditional sites to mate⁵. If so, analysis of the current state of the environment may yield incorrect predictions about mating sites^{5–9}. Mating sites of the bluehead wrasse *Thalassoma bifasciatum* have remained in daily use over 12 years (four generations) without changing locations. Here, I show that experimental replacement of entire local populations led to the use of new sites, which continued to be used after the manipulations. Thus mating site locations may not be the result of individual assessment of current resource quality, but instead represent culturally transmitted traditions. Attempts to explain mating patterns based on the distribution of resources should therefore be approached with caution.

The bluehead wrasse is a common Caribbean coral reef fish that mates daily throughout the year¹⁰. It was studied in small patch reefs (up to 1,500 m²) in the San Blas Islands off Panama¹¹. On reefs of this size, females migrate to specific spawning sites, each of which are defended by a single large brightly-coloured male. These males do not seem to exercise mate choice within their territories (E. van den Berghe and R.R.W., preprint) nor is female choice of mating site strongly influenced by the male present¹². There is no parental care; spawning consists of egg and sperm release at the apex of a rapid swim towards the surface. The mixing and transport which occurs during the long pelagic larval period of this species (average of 52 days¹³) strongly suggests that reef populations are not subject to long-term local genetic differentiation. The spawning sites themselves are located on shallow upward projections near the reef periphery and are concentrated at the downcurrent ends of reefs (Fig. 1). These characteristics presumably help reduce reef-based egg predation¹⁴.

Over the past 12 years the exact locations of spawning sites have been monitored on 22 patch reefs in the San Blas field site. Although the mating success associated with a particular site can change greatly with time¹², the actual site locations have remained strikingly constant. In spite of major fluctuations in population sizes, no new sites have come into use, nor has an old site ever been lost ($N = 87$ sites). The bluehead wrasse lives no more than three years^{13,14}, so these observations span about four generations.

This constancy of site usage is remarkable in light of the 3–7-fold excess of sites that fit the physical criteria shown in Fig. 1 over those that are actually used (Table 1). These observations do not eliminate the possibility that individuals are continually monitoring suitable sites and making optimal choices based on some unmeasured cues.

To differentiate properly between resource assessment and traditionality, I translocated entire populations to previously studied reefs from which all adult residents had been removed and then monitored subsequent site use. If resource assessment controls mating-site use, the newly transplanted populations would be expected to use the previous mating sites. Traditionality, in contrast, can constrain behaviour to particular sites even if the underlying resource base has changed^{5–7}. Traditionality would be indicated if new sites came into use and persisted long

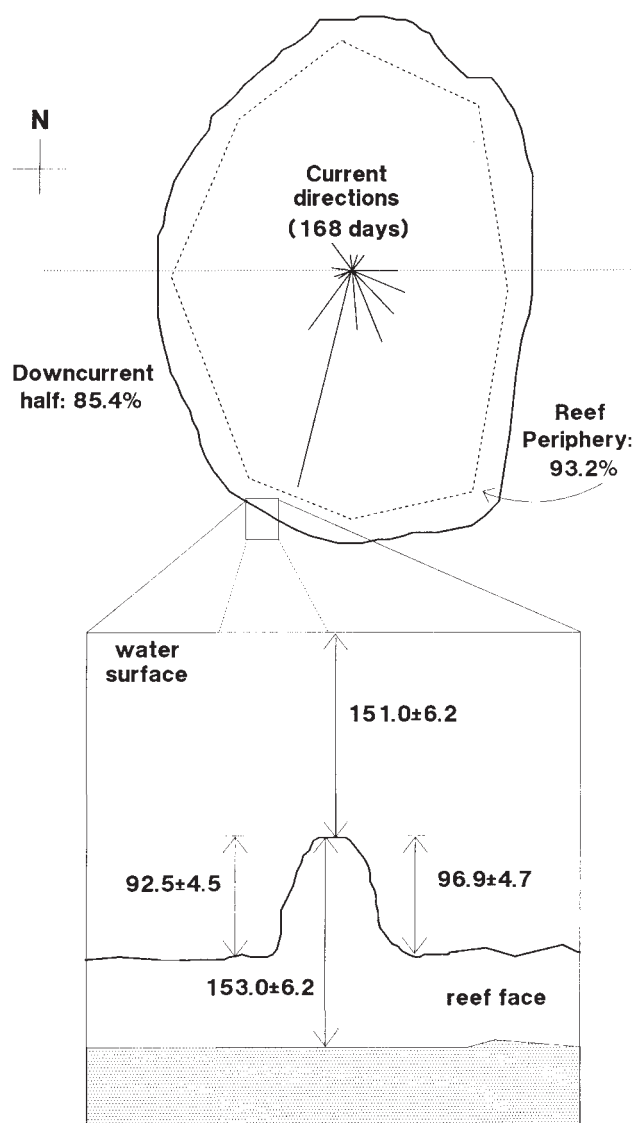


Fig. 1 Depiction of the physical characteristics of spawning sites used by the bluehead wrasse, compiled from measurements of 103 sites on 25 reefs. Top, site distribution on a composite reef as a percentage of the total sites sampled. Daily current direction data for 168 days are superimposed on the reef outline; mean ($\pm$ one angular deviation) current direction was $172.8^\circ \pm 49.5^\circ$. Reef periphery was designated as all areas < 2 m from the reef edge. Measurements (cm) of mean ($\pm$ one standard error) site depth and vertical projection from the reef edge are shown in the lower part of the figure.

Table 1 Influence of pre-manipulation site usage on the location of mating sites

	Total sites available	Sites used before	Sites used after	Sites used before and after	Contingency table analysis
Experimental reefs					
PV20 (1983)	24	6	6	2	$P = 0.48$
PT32 (1984)	35	10	10	5	$P = 0.09$
WI26 (1983)	29	7	8	4	$P = 0.07$
PT31 (1984)	43	6	7	2	$P = 0.25$
PV8 (1984)	34	7	13	5	$P = 0.06$
PV21 (1986)	40	7	6	1	$P = 0.71$
Control reefs					
PV19 (1983)	43	8	8	8	$P < 0.001$
PV21 (1984)	40	7	7	7	$P < 0.001$
PV9 (1984)	45	12	12	12	$P < 0.001$

Residents were replaced with naive populations. The null hypothesis of the analysis is that the sites used after the manipulation represent a random draw from the available pool, with no influence of past usage. Probabilities are given for a one-tailed Fisher's Exact Test. Potential mating sites were upward projections (measurements within two standard errors of the mean shown in Fig. 1) on the downcurrent periphery of a reef. All removals and transfers were done with a lift net baited with sea urchins. Complete removal took two days on average and the reef was subsequently monitored for an additional two days to ensure that all *T. bifasciatum* had been caught. The total number of individuals removed ranged from 123 to 422 and these were transferred to distant (> 2 km) reefs otherwise unused for this study. Replacement males and females came from separate reefs; the number of individuals transferred was equal to the number of individuals removed and the sex ratio was maintained. Tagging studies have shown that individuals do not tend to move naturally between reefs²⁶. Monitoring of mating sites before and after the manipulation was performed over at least 15 mating periods each by divers positioned along the reef periphery. Such monitoring continued until a minimum of 200 matings had been recorded. All reefs were surveyed for post-manipulation mating site usage during the following periods: July–August 1983; October 1983; June–August 1984; April–May 1985; July–August 1985; June–August 1986; July 1987. During these surveys, all downcurrent portions of the reefs were monitored for reproductive activity during the mating period for at least five days.

new sites had been identified as potential sites before the start of the experiment.

Two of the reefs were manipulated as early as 1983 and all have been repeatedly monitored to track continued site usage since the time of the experiments. The sites which were used after the manipulation continued to be used up until the last survey, in June 1987. As in the pre-manipulation survey, no new sites spontaneously appeared.

Thus, under natural circumstances, mating site locations do not seem to be determined by individual assessment of current resources, but instead through cultural transmission of traditional sites. Once tradition is broken, other sites may be used and they in turn become traditional centres of activity used over generations. Although traditional resting places and migration routes have been shown to exist for another coral reef fish¹⁵, this is the first experimental demonstration that traditionality determines mating-site use.

Traditional behaviour can bypass a potentially dangerous time of trial-and-error learning and is especially beneficial when naive young disperse into unknown habitats^{7–9}. Cultural transmission should not be difficult to achieve in this species, because reproduction and recruitment occur throughout the year^{13,16}. Females seem to learn mating-site locations from experienced females. Transplanted females mate only at established sites if native females are present, but will establish new sites if only native males are present (R.R.W., manuscript in preparation).

after the manipulation. Controls involved removing and subsequently replacing entire populations on their home reefs.

I performed whole population replacements on six reefs and used three other reefs as controls. In every experimental replacement, the results were consistent with the traditionality hypothesis: sites that came into use after the manipulation represented a random sample of all potential sites (Table 1). That is, sites used previously were no more likely to be used than some other site that fitted the physical criteria shown in Fig. 1. Control populations, in contrast, re-used all the same sites (Table 1). Overall, the null hypothesis that experiments and controls have the same effect on site re-use can be rejected (Mann–Whitney *U*-test of proportion of sites re-used, $U_s = 18$, $P < 0.05$). The number of sites used before and after the experimental manipulations was quite similar in most cases, and all

Mating resource assessment behaviour exists in many species¹⁷⁻²⁰. Bluehead wrasses may not exercise such assessment because there may be few or no fitness differences among potential sites²¹. In this case, the time saved and safety gained through traditional behaviour may well outweigh any benefits derived from frequent careful assessment²². However, traditions can continue in the same area after a resource base is altered because they are, by definition, less responsive to external conditions⁵⁻⁷. This study should inject a note of caution into attempts to test hypotheses about the evolution of mating systems by

measuring the current spatial distribution of resources or females²³⁻²⁵.

I thank T. Boundy, T. Brown, K. Clifton, L. Clifton, D. Wiley, J. Crisp, L. Fore, P. Hart, J. Jolly, J. and V. Maldonado, G. McKinley, B. McAfee, S. Olivier, M. Schildhauer and E. Schultz for field assistance. K. Clifton, E. van den Berghe, J. Endler, S. Rothstein and E. Schultz read the manuscript. The Kuna people of the Comarca de San Blas, Panama, kindly provided permission to study in their water. This work was supported by the NSF.

Received 16 May; accepted 1 September 1988.

1. Parker, G. A. in *Mate Choice* (ed. Bateson, P.) 141-166 (Cambridge University Press, Cambridge, 1983).
2. Partridge, L. & Halliday, T. R. in *Behavioural Ecology, an Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 222-250 (Sinauer, Sunderland, 1984).
3. Emlen, S. T. & Oring, L. W. *Science* **197**, 215-223 (1977).
4. Vehrencamp, S. L. & Bradbury, J. W. in *Behavioural Ecology, an Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 251-278 (Sinauer, Sunderland, 1984).
5. Parker, G. A. *Rev. Ent.* **23**, 173-196 (1978).
6. Bonner, J. T. *The Evolution of Culture in Animals* (Princeton University Press, 1980).
7. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (University of Chicago Press, 1985).
8. Galef, B. G. Jr. *Adv. Study Behav.* **6**, 77-100 (1976).
9. Mainardi, D. in *Sociobiology: Beyond Nature/Nurture?* (eds Barlow, G. W. & Silverberg, J.) 227-255 (Westview, Boulder, 1980).
10. Warner, R. R. & Robertson, D. R. *Smithson. Contr. Zool.* **254**, 1-27 (1978).

11. Robertson, D. R. *Copeia* **1987**, 637-642 (1987).
12. Warner, R. R. *Anim. Behav.* **35**, 1470-1478 (1987).
13. Victor, B. C. *Ecol. Monogr.* **56**, 145-160 (1986).
14. Randall, J. E. & Randall, H. A. *Zoologica (N.Y.)* **48**, 49-60 (1963).
15. Helfman, G. S. & Schultz, E. T. *Anim. Behav.* **32**, 379-384 (1984).
16. Warner, R. R. *Evolution* **38**, 148-162 (1984).
17. Searcy, W. A. *Am. Nat.* **114**, 77-100 (1979).
18. Pleszczyńska, W. K. & Hansell, R. I. C. *Am. Nat.* **116**, 821-830 (1980).
19. Borgia, G. *Evolution* **36**, 307-315 (1982).
20. Thornhill, R. & Alcock, J. *The Evolution of Insect Mating Systems* (Harvard University Press, 1983).
21. Shapiro, D. Y., Hensley, D. A. & Appeldoorn, R. S. *Envir. Biol. Fish.* **22**, 3-14 (1988).
22. Warner, R. R. *Proc. Fifth Int. Coral Reef Congress* **4**, 145-150 (1985).
23. Lill, A. Z. *Tierpsychol.* **36**, 513-530 (1974).
24. Robertson, D. R. & Hoffman, S. G. *Z. Tierpsychol.* **45**, 298-320 (1977).
25. Bradbury, J. W., Gibson, R. & Tsai, I. M. *Anim. Behav.* **34**, 1694-1709 (1986).
26. Warner, R. R. & Hoffmann, S. G. *Evolution* **34**, 508-518.

Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms

Andrew H. Paterson*, Eric S. Lander†‡, John D. Hewitt§, Susan Peterson*, Stephen E. Lincoln† & Steven D. Tanksley*

* Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853, USA

† Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

‡ Harvard University, Cambridge, Massachusetts 02138, USA

§ Department of Vegetable Crops, University of California, Davis, California 95616, USA

The conflict between the Mendelian theory of particulate inheritance¹ and the observation of continuous variation for most traits in nature was resolved in the early 1900s by the concept that quantitative traits can result from segregation of multiple genes, modified by environmental effects²⁻⁵. Although pioneering experiments⁶⁻⁹ showed that linkage could occasionally be detected to such quantitative trait loci (QTLs), accurate and systematic mapping of QTLs has not been possible because the inheritance of an entire genome could not be studied with genetic markers⁷. The use of restriction fragment length polymorphisms¹⁰ (RFLPs) has made such investigations possible, at least in principle. Here, we report the first use of a complete RFLP linkage map to resolve quantitative traits into discrete Mendelian factors, in an interspecific back-cross of tomato. Applying new analytical methods, we mapped at least six QTLs controlling fruit mass, four QTLs for the concentration of soluble solids and five QTLs for fruit pH. This approach is broadly applicable to the genetic dissection of quantitative inheritance of physiological, morphological and behavioural traits in any higher plant or animal.

The parents for the back-cross were the domestic tomato *Lycopersicon esculentum* cv. UC82B (denoted E) and a wild South American green-fruited tomato *L. chmielewskii*¹¹ accession LA1028 (denoted CL). These strains have very different fruit masses (E ~ 65 g; CL ~ 5 g) and concentrations of soluble solids¹² (E ~ 5%; CL ~ 10%)—traits of agricultural

importance, because they jointly determine the yield of tomato paste. In addition, the strains are known to be polymorphic for genes affecting fruit pH, which is important for the optimal preservation of tomato products¹³; the difference in pH between the parental strains is, however, small.

A total of 237 back-cross plants, with E as the recurrent parent, were grown in the field at Davis, California. Between five and 20 fruit from each plant were assayed¹³ for fruit mass, soluble-solids concentration (°Brix; see Fig. 1 legend for definition) and pH, each of which showed continuous variation (Fig. 1). Soluble-solids concentration correlated negatively with fruit mass ($r = -0.42$) and positively with pH ($r = +0.33$).

We had previously constructed a genetic linkage map of tomato¹⁴ with over 300 RFLPs and 20 isozyme markers, by analysing 46 F₂ individuals derived from *L. esculentum* cv. VF36 × *L. pennellii* accession LA716 (E × P). The map is essentially complete: it has linkage groups covering all 12 tomato chromosomes with an average spacing of 5 cM between markers (1 cM is the distance along the chromosome which gives a recombination frequency of one per cent). For QTL mapping, we selected a subset of markers spaced at approximately 20 cM intervals and displaying polymorphism between the E and CL strains. These included 63 RFLPs and five isozyme markers. In addition, the E and CL strains differ in two easily-scored, simply-inherited morphological traits: determinacy (described below) and uniform ripening, controlled by the *sp* and *u* genes, respectively. Although a few distal regions did not contain appropriate markers, we estimate that about 95% of the tomato genome was detectably linked to the markers used.

These 70 genetic markers were scored for each of the 237 E × CL back-cross progeny (as described in ref. 13), and a linkage map was constructed *de novo* using MAPMAKER¹⁵. The map covers all 12 chromosomes with an average spacing of 14.3 cM. Although the linear order of markers inferred from the E × CL cross essentially agreed with that inferred from the E × P cross (but see Fig. 3 legend), genetic distances differed markedly in certain intervals (for example, 51 cM in E × P and 11 cM in E × CL, for the distance between the 45S ribosomal repeat and *TG1B* on chromosome 2). In total, the markers scored in both crosses span 852 cM in the E × CL map versus 1103 cM in the E × P map, a highly significant ($P < 0.01$) difference. Skewed segregation ($P < 0.05$) was detected for 48 of the 70 markers, comprising 21 distinct regions distributed over all 12 chromosomes. The heterozygote (E/CL) was over-

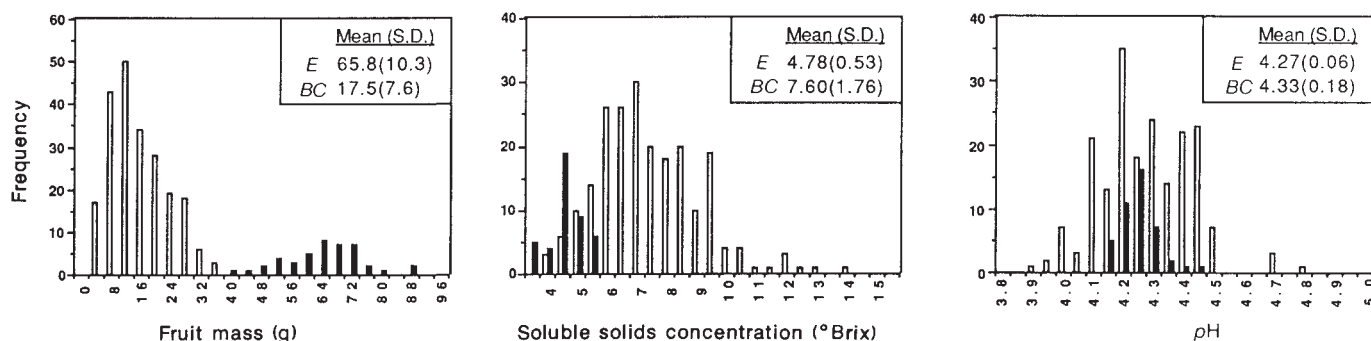
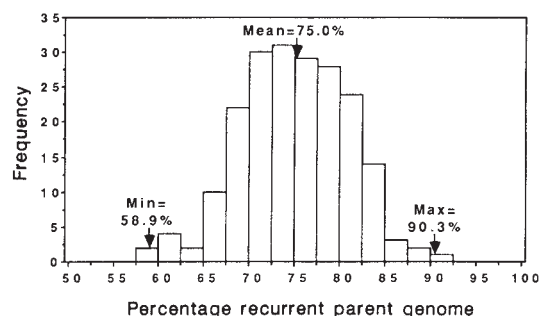


Fig. 1 Frequency distribution for fruit mass, soluble-solids concentration (°Brix, a standard refractometric measure primarily detecting reducing sugars, but also affected by other soluble constituents; 1 °Brix is approximately 1% w/w) and pH in the E parental strain and in the back-cross (BC) progeny. The tomatoes were grown in the field at Davis, California, in a completely randomized design including 237 BC plants (with E as the recurrent pistillate parent), as well as E, CL and the F_1 as controls. Neither CL nor the F_1 progeny matured completely, as is typical in the central valley of California. Among the BC plants, six failed to mature and 12 produced too few fruit to assay reliably for quantitative traits. The absence of quantitative trait data for these few progeny should yield at most a slight bias in our analyses. Means and standard deviations for the distributions of the E parental strain (E filled bars) and the BC progeny (BC open bars) appear in the upper right of each histogram. The distributions for soluble-solids concentration and pH are approximately normal. The distribution of the BC progeny for fruit weight is clearly skewed; $\log_{10}$ (fruit mass) was studied throughout to achieve approximate normality (see ref. 5; $E = 1.81 \pm 0.07$; $BC = 1.20 \pm 0.19$). The proportion of variance due to environment was estimated to be the square of the ratio of the standard deviations (E/BC), for log-mass, solids and pH.

Fig. 2 Distribution of percentage of recurrent parent (E) genotype in the 237 back-cross progeny, estimated on the basis of the marker genotypes and their relative distances. Determination of marker genotypes was as previously described¹³. Estimates of the percentage of recurrent parent genome were produced by the recently developed computer program HyperGene™ (N. D. Young, A.H.P. and S.D.T., unpublished results). Although the average agreed closely with the Mendelian expectation of 75% for a back-cross, values for individual plants ranged from 59% to over 90%. The distribution of the proportion of recurrent-parent genome agrees with the mathematical expectation^{35,36}. The individual with >90% E appears to carry only five fragments from CL (ranging from 9 to 47 map units in length) and could be returned to essentially 100% E with two additional back-crosses of fewer than 100 plants each, or one additional back-cross of about 550 plants. This is far more rapid than the 6–8 back-crosses routinely used to eliminate donor genome in the absence of markers.



abundant in 12 cases, whereas in nine cases the homozygote (E/E) was favoured. Overall, the effects of skewing approximately cancelled each other out: on average, the back-cross contained the expected 75% E genome (Fig. 2).

We then turned to the question of mapping the Mendelian factors that underly continuous variation in fruit mass, soluble-solids concentration and pH. The method of maximum likelihood and lod scores, commonly used in human linkage analysis¹⁷, has recently been adapted¹⁶ to allow interval mapping of QTLs. At each position in the genome, one computes the 'most likely' phenotypic effect of a putative QTL affecting a trait (the effect which maximizes the likelihood of the observed data arising) and the odds ratio (the chance that the data would arise from a QTL with this effect divided by the chance that it would arise given no linked QTL). The lod score, defined as the $\log_{10}$ of the odds ratio, summarizes the strength of evidence in favour of the existence of a QTL with this effect at this position; if the lod score exceeds a pre-determined threshold, the presence of a QTL is inferred. The traditional approach^{8,9} to mapping QTLs involves standard linear regression, which accurately measures the effect of QTLs falling at marker loci only, underestimating the effects of other loci in proportion to the amount of recombination between marker and QTL. In contrast, interval mapping allows inference about points throughout the entire genome and

avoids confounding phenotypic effects with recombination, by using information from flanking genetic markers. In the special case when a QTL falls exactly at a marker locus, interval mapping reduces to linear regression. A computer program, MAPMAKER-QTL, was written (S.E.L. and E.S.L., unpublished) to implement interval mapping.

Due to the large number of markers tested, an extremely high lod score threshold must be adopted to avoid false positives. Given the genetic length of the tomato genome and the density of markers used, a threshold of 2.4 gives a probability of less than 5% that even a single false positive will occur anywhere in the genome¹⁶. This is approximately equivalent to requiring the significance level for any single test to be 0.001.

QTL likelihood maps, showing how lod scores for fruit mass, soluble-solids concentration and pH change as one moves along the genome, reveal multiple QTLs for each trait and estimate their location to within 20–30 cM (Fig. 3).

(1) Factors for fruit mass were found on six chromosomes (1, 4, 6, 7, 9 and 11). In each case, CL alleles decrease fruit mass (by 3.5 to 6.0 g), adding to a total reduction of 28.1 g inferred for back-cross progeny carrying a CL allele at all six loci. This accounts for about half of the approximately 60 g difference between E and CL.

(2) Factors for soluble-solids concentration were found on

four chromosomes (3, 4, 6 and 7). In each case, CL alleles elevate soluble-solids concentration (by 0.83 to 1.89 °Brix), adding to a total of 4.57 °Brix (versus a difference of ~5 °Brix between the parental strains). This large effect in the back-cross is consistent with previous reports that high soluble-solids concentration exhibits dominance¹² and overdominance¹³. The QTL alleles for both fruit mass and soluble-solids concentration all produce effects in the direction predicted by the difference between the parental strains.

(3) Factors for pH were found on five chromosomes (3, 6, 7, 8 and 10). In addition, the lod score for a putative QTL on chromosome 9 fell just below our threshold. Because the parental strains do not differ greatly in pH, we suspected that CL alleles might not all produce effects in the same direction. In fact, pH was increased by four QTLs and decreased by two, including the likely QTL on chromosome 9. This provides a genetic explanation for the observation that many back-cross progeny exhibited more extreme phenotypes than the parental strains (Fig. 1), a phenomenon known as transgression¹⁸.

Together, the QTLs identified for fruit mass, soluble solids and pH account for 58%, 44% and 48%, respectively, of the phenotypic variance among the back-cross progeny, with another 13%, 9% and 11% attributable to environment.

The numbers of QTLs reported for each trait must be considered a minimum estimate. Because an extremely stringent threshold was used to avoid any false positives, some sub-threshold effects probably represent real QTLs. For example, the regions near *TG19* on chromosome 1, *CD41* on chromosome 5 and *TG68* on chromosome 12 may affect soluble-solids concentration and merit further attention in larger populations. Similarly, the region near the *u* locus on chromosome 10 may contain an additional QTL affecting pH (see Fig. 3 legend). Moreover, we cannot rule out the presence of many additional QTLs with tiny phenotypic effects—postulated in evolutionary theory¹⁹ and supported by some experimental evidence²⁰. Also, it is conceivable that some of our apparent QTLs actually represent several closely-linked QTLs, each with small phenotypic effects in the same direction—a phenomenon that might arise particularly in regions of genetic map compression. Finally, we should emphasize that the QTL mapping here strictly applies only to the specific environment tested and to heterozygosity for CL alleles. In principle, homozygosity for CL alleles could have been studied by using an F₂ self between E and CL, but in practice too many of the progeny are sterile.

Some regions of the genome clearly exert effects on more than one trait (for example, chromosome 6; Fig. 3), providing a genetic explanation for at least some of the correlation between the traits. Although the present data are insufficient to distinguish between pleiotropic effects of a single gene and independent effects of tightly-linked loci, the frequent coincidence of QTL locations for different traits makes it likely that at least some of the effects are due to pleiotropy.

The region near *sp* on chromosome 6 has the largest effects on soluble solids and pH, as well as a substantial effect on fruit mass. The *sp* gene affects plant-growth habit: the dominant CL allele causes continuous apical growth (indeterminate habit), whereas the recessive E allele causes termination in an inflorescence ('determinate' or 'self-pruning' habit)²¹. Although indeterminacy has been reported previously²² to elevate both fruit mass and soluble-solids concentration within *L. esculentum*, we associated it with reduced fruit mass in both E × CL and another interspecific cross (E × *L. cheesmanii*; A.H.P., S. Damon, J.D.H. and S.D.T., unpublished data). These differing results might be due to a second, tightly-linked locus or to unlinked modifier genes.

Overall, pairwise epistatic interactions between intervals were not common (about 5% of two-way analysis-of-variance tests were significant at 0.05). An interesting exception was the region near *TG16* on chromosome 8, at which the CL allele significantly enhanced the effect of three of the four QTLs for soluble-solids

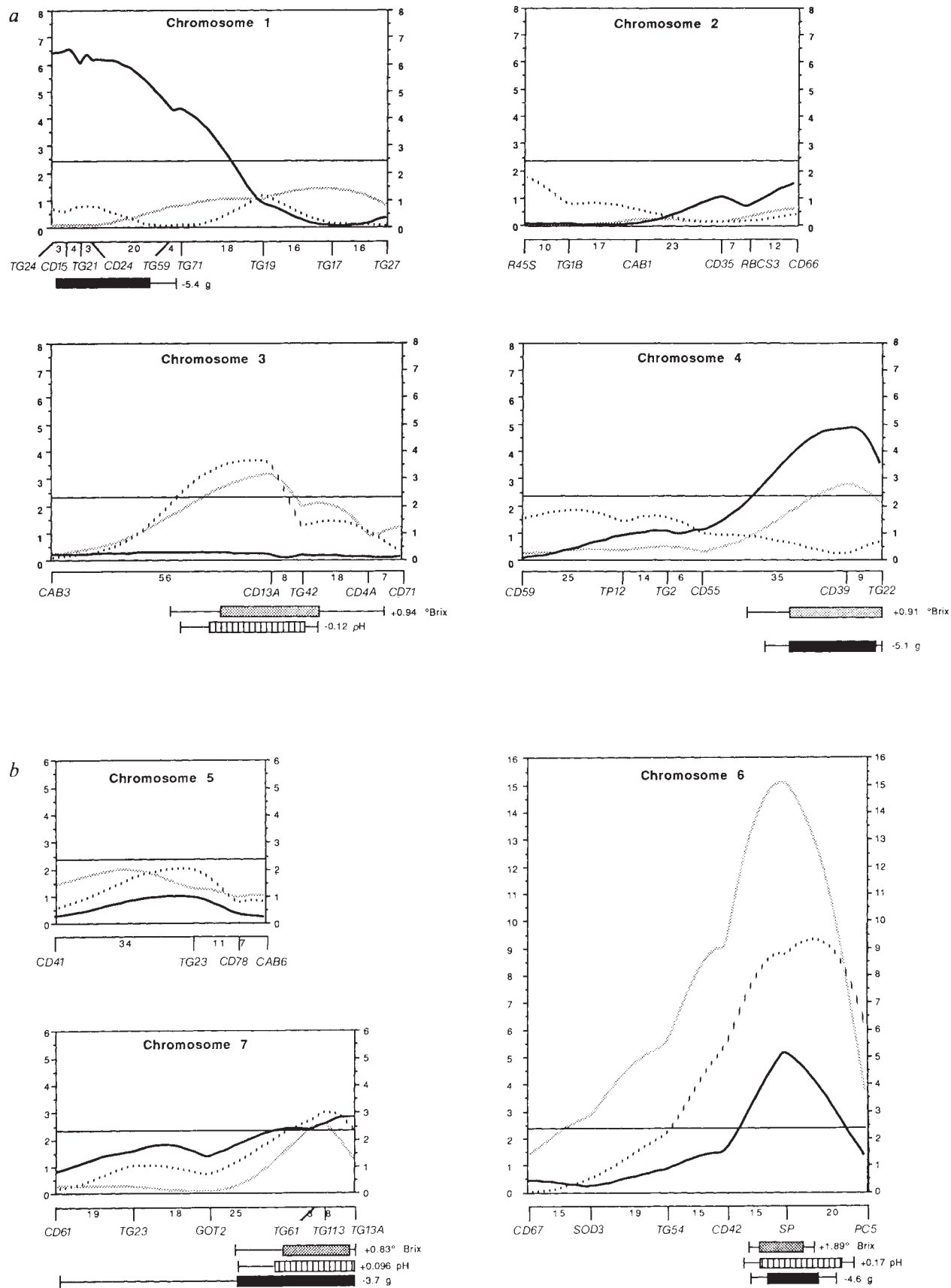
concentration. *TG16* also showed the most extreme segregation distortion of any marker scored (about 4:1 in favour of the E/E homozygote) and is in a region known to exhibit skewed segregation in back-crosses to other green-fruited tomato species^{23,24}. The unusual properties of this region of CL clearly merit further study.

The QTLs identified here may well differ from those that would be fixed by repeated back-crossing with continuing selection for a trait, a classical method for introgressing quantitative traits. Work on LA1563, a strain with increased soluble solids produced¹² through back-crossing a different strain of E to CL has provided some suggestive evidence. By surveying RFLPs, Tanksley and Hewitt¹³ recently found that LA1563 has maintained three separate regions from CL: near *CD56* on chromosome 10, near *Got2* on chromosome 7 and near *TG13* on chromosome 7. Here, we detected above-threshold effects in the last of these three regions only (which, interestingly, failed to show effects on soluble solids in a single-environment test by Tanksley and Hewitt¹³). Moreover, we detected QTLs affecting soluble-solids concentration in regions that did not seem to be retained. Unfortunately, the results of the two experiments are not directly comparable due to the use of a different E strain by Rick, possible environmental differences between the experiments, the possibility that small CL fragments containing QTLs went undetected in LA1563, the possibility that the region near *TG13* retained in LA1563 may not contain the QTL we detected here and the possibility that some of our sub-threshold effects are real. Although more detailed studies are clearly needed, it is interesting to speculate about why repeated back-crossing may fix a narrower class of QTLs than found by QTL mapping. Because such breeding programs¹² demand horticultural acceptability, they are likely to select against otherwise-desirable QTLs which are closely linked to undesirable effects from the wild parent. If such QTLs can first be identified by mapping, it may be feasible to remove linked deleterious effects by recombination.

Having mapped several QTLs with relatively large effects, we are now making crosses to isolate them in near-isogenic lines. These lines will be used to characterize the QTLs in various dosages, genetic backgrounds, environments and combinations. By re-assembling selected CL alleles in an otherwise E genotype, we hope to engineer an agriculturally-useful tomato with a higher yield of soluble solids.

The general approach of QTL mapping is broadly applicable to a wide range of biological endeavours. In agriculture, it might be desirable to transfer to domestic strains many quantitative traits harboured in wild species, including resistance to diseases and pests, tolerance to drought, heat, cold and other adverse conditions, efficient use of resources and high nutritional quality^{25,26}. In mammalian physiology, selective breeding has generated rodent strains which differ greatly in quantitative traits such as hypertension, atherosclerosis, diabetes, predispositions to cancer, drug sensitivities and various behavioural patterns; information on the number, location and nature of these QTLs would be of value in medicine^{16,27}. In evolutionary biology, the process of speciation can be investigated by studying the number and nature of genes underlying reproductive isolation²⁸.

The availability of detailed RFLP linkage maps^{14,29-34} makes it possible to dissect quantitative traits into discrete genetic factors (QTLs): all regions of a genome can be assayed and accurate estimates of phenotypic effects and genetic position derived from interval analysis¹⁶. Once QTLs are mapped, RFLP markers permit genetic manipulations such as rapid construction of near-isogenic lines: flanking markers may be used to retain the QTL and the study of the remaining markers may be used to speed progress by identifying individuals with a fortuitously high proportion of the desired genetic background³⁸ (Fig. 2 legend). Using isogenic lines, the fundamental tools of genetics and molecular biology may be brought to bear on the study of QTLs—including testing of complementation, dominance and



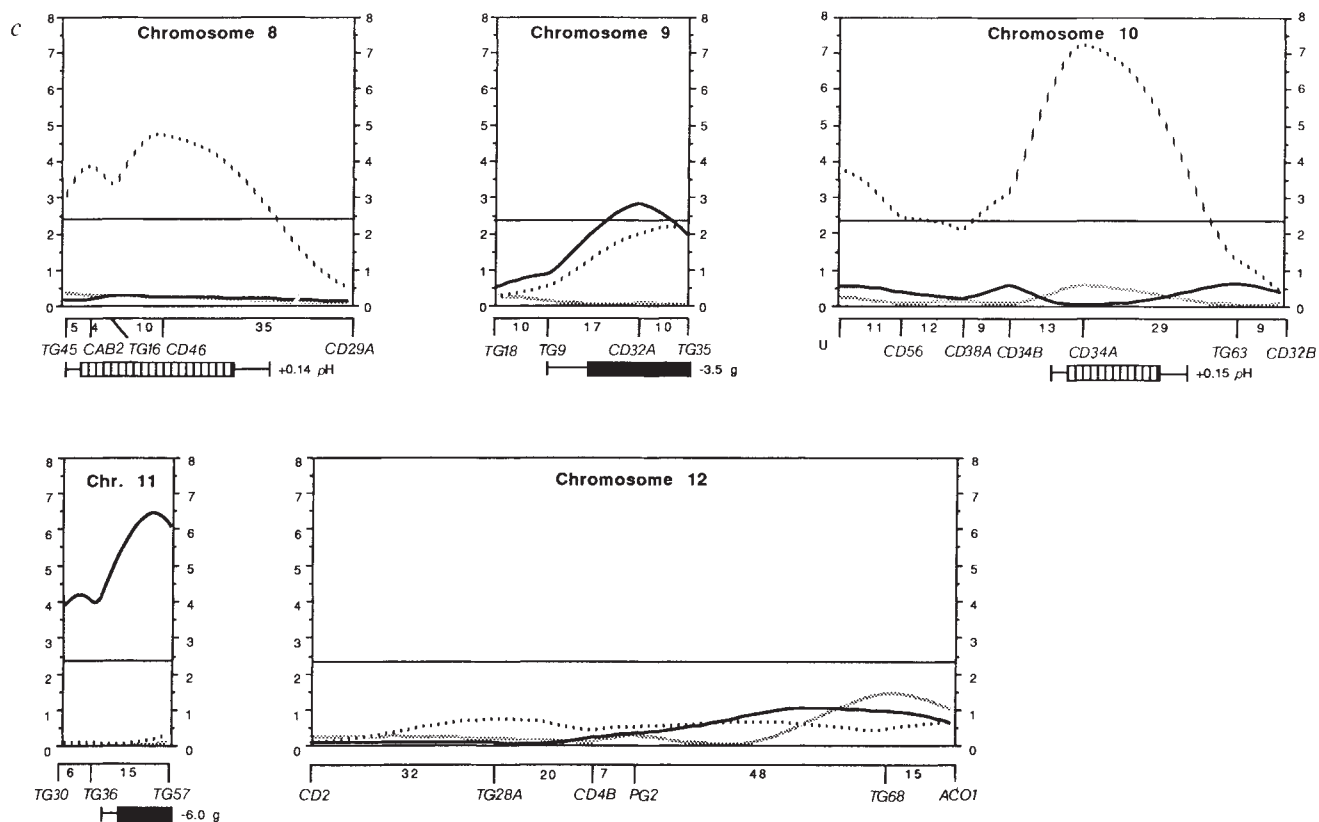


Fig. 3 QTL likelihood maps indicating lod scores for fruit mass (solid lines and bars), soluble-solids concentration (dotted lines and bars) and pH (hatched lines and bars), throughout the 862 cM spanned by the 70 genetic markers. The RFLP linkage map used in the analysis is presented along the abscissa, in Kosambi³⁷ cM. (The order of the markers agrees with the previously published map¹⁴ of the E×P cross, except for three inversions of adjacent markers: (TG24-CD15), (TG63-CD32B) and (TG30-TG36). In the first case, re-analysis of the E×P data with MAPMAKER¹⁵ indicates that the order shown here is the more likely in both E×P and E×CL. For the other two, the orders shown here are more likely in E×CL by odds of 10⁴:1 and 10⁷:1, but the inverse is more likely in E×P by 11:1 and 8:1 odds. These differences will be investigated in a larger E×P population.) Soluble-solids concentration and pH were analysed in °Brix and pH units, respectively; allele effects on fruit mass are presented in g; log-transformation of fruit mass was used in all analyses to achieve approximate normality. The maximum likelihood effect of a putative QTL, as well as the lod score in favour of the existence of such a QTL, have been determined at points spaced every 1 cM throughout the genome, according to Lander and Botstein¹⁶ and a smooth curve plotted through the points. The height of the curve indicates the strength of the evidence (log₁₀ of the odds ratio) for the presence of a QTL at each location—not the magnitude of the inferred allelic effect. The horizontal line at a height of 2.4 indicates the stringent threshold that the lod score must cross to allow the presence of a QTL to be inferred (see text). Information about the likely position of the QTL can also be inferred from the curve. The maximum likelihood position of the QTL is the highest point on the curve. Bars below each graph indicate a 10:1 likelihood support interval¹⁶ for the position of the QTL (the range outside which the likelihood falls by a lod score of 1.0), whereas the lines extending out from the bars indicate a 100:1 support interval. Phenotypic effects indicated beside the bars are the inferred effect of substituting a single CL allele for one of the two E alleles at the QTL. Several regions show sub-threshold effects on one or more traits (chromosome one near TG19, chromosome five near TG32 and chromosome 12 near TG68) which may represent QTLs but this requires additional testing. The region near TG68 may be particularly interesting, as it is the only instance found where the CL allele seems to decrease soluble-solids concentration (by about 0.7 °Brix). In the case of chromosome 10, the lod score for pH crosses the significance threshold in two places. Controlling for the presence of a QTL near CD34A, we tested for the presence of a second QTL near u (by comparing the maximum lod scores assuming the presence of only the first QTL to the maximum lod score assuming the presence of two QTLs). Allowing for a QTL in the region of CD34A, the residual lod score near u falls below the required threshold. Thus, the evidence is not yet sufficient to support the presence of a QTL near u.

Methods. The lod score and the maximum likelihood estimate (MLE) of the phenotypic effect at any point in the genome is computed assuming that the distribution of phenotypes in the BC progeny represents a mixture of two normal distributions (of equal variance) with means depending on the genotype at a putative QTL at the given position. (Note that QTLs are considered individually and thus we did not assume that different QTL effects can be added—except in studying the possibility of two QTLs on chromosome 10 affecting pH.) Specifically, at a given position in the genome, the likelihood function for individual *i* with quantitative phenotype ϕ_i is given by $L_i(\alpha, \sigma) = (2\pi\sigma^2)^{-1/2} \{p_1 \exp(-\phi_i^2/2\sigma^2) + p_2 \exp(-(\phi_i - \alpha)^2/2\sigma^2)\}$, where α is the effect of substituting a CL allele for an E allele at a putative QTL in the given position, σ^2 is the phenotypic variance not attributable to the QTL and p_1 and p_2 are the probabilities that individual *i* has genotype E/E and E/CL, respectively, at the QTL (which can be computed on the basis of the genotypes at the flanking markers and the distance to the flanking markers). The likelihood function for the entire population is $L = \prod L_i$. Also α^* and σ^* denote the MLEs allowing the possibility of a QTL at the location (the values which maximize L) and σ^{**} denotes the MLE of σ , subject to the constraint that no QTL is linked ($\alpha = 0$). The lod score is then given by $\log_{10}\{L(\alpha^*, \sigma^*)/L(0, \sigma^{**})\}$. This method for QTL mapping is developed more fully in ref. 16.

epistasis; characterization of physiological and biochemical differences between isogenic lines; isolation of additional alleles by mutagenesis (at least in favourable systems); and, eventually, physical mapping and molecular cloning of genetic factors underlying quantitative traits.

We thank Janice Chen, Mark Daly, Gerald Dickinson and Mitzi Aguirre for technical assistance. We thank our colleagues and the referees for comments on the manuscript. This work was supported in part by grants from the NSF (to E.S.L. and to S.D.T.), from the System Development Foundation (to E.S.L.), and from the US Department of Agriculture (to S.D.T.).

Received 8 July; accepted 9 September 1988.

- Mendel, G. *Verh. naturf. Ver. in Brunn*, **4**, (1868).
- Johannsen, W. *Elemente der exakten Erblchkeitslehre* (Fischer, Jena, 1909).
- Nilsson-Ehle, H. *Kreuzungsuntersuchungen an Hafer und Weizen* (Lund, 1909).
- East, E. M. *Genetics*, **1**, 164-176 (1915).
- Wright, S. *Evolution and the Genetics of Populations* (University of Chicago Press, 1968).
- Sax, K. *Genetics*, **8**, 552-556 (1923).
- Thoday, J. M. *Nature*, **191**, 368-369 (1961).
- Tanksley, S. D., Medina-Filho, H. & Rick, C. *Heredity*, **49**, 11-25 (1982).
- Edwards, M. D., Stuber, C. W. & Wendel, J. F. *Genetics*, **116**, 113-125 (1987).
- Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.*, **32**, 314-331 (1980).
- Chmielewski, T. *Genet. pol.*, **9**, 97-124 (1968).
- Rick, C. M. *Hilgardia*, **42**, 493-510 (1974).
- Tanksley, S. D. & Hewitt, J. *Theor. appl. Genet.*, **75**, 811-823 (1988).
- Tanksley, S. D., Miller, J., Paterson, A. & Bernatzky, R. *Proc. 18th Stadler Genet. Symp.* (in the press).
- Lander, E. S. *et al. Genomics*, **1**, 174-181 (1987).
- Lander, E. S. & Botstein, D. *Genetics* (in the press).
- Ott, J. *Analysis of Human Genetic Linkage* (Johns Hopkins, Baltimore, 1985).
- Simmonds, N. W. *Principles of Crop Improvement*, 82-85 (Longman, New York, 1981).
- Lande, R. *Heredity*, **50**, 47-65 (1983).
- Shrimpton, A. E. & Robertson, A. *Genetics*, **11**, 445-459 (1988).
- Yeager, A. F. *J. Hered.*, **18**, 263-265 (1927).
- Emery, G. C. & Munger, H. M. *J. Am. Soc. Hort. Sci.*, **95**, 410-412 (1966).
- Zamir, D. & Tadmor, Y. *Bot. Gaz.*, **147**, 355-358 (1986).
- Tanksley, S. D. in *Isozymes in Plant Genetics and Breeding* (eds Tanksley, S. D. & Orton, T. J.) 331-338 (Elsevier, Amsterdam, 1983).
- Rick, C. M. in *Genes, Enzymes and Populations* (ed. A. M. Srb) 255-268 (Plenum, New York, 1973).
- Harlan, J. R. *Crop Sci.*, **16**, 329-333 (1976).
- Festing, M. F. W. *Inbred Strains in Biomedical Research* (Oxford, New York, 1979).
- Coyne, J. A. & Charlesworth, B. *Heredity*, **57**, 243-246 (1986).
- Helentjaris, T. *Trends Genet.*, **3**, 217-221 (1987).
- Landry, B. S., Kessel, R., Leung, H. & Michelsmore, R. W. *Theor. appl. Genet.*, **74**, 646-653 (1987).
- Burr, B., Burr, F. A., Thompson, K. H., Albertsen, M. C. & Stuber, C. W. *Genetics*, **118**, 519-526 (1988).
- Chang, C., Bowman, J. L., DeJohn, A. W., Lander, E. S. & Meyerowitz, E. M. *Proc. natn. Acad. Sci. U.S.A.*, **68**, 6856-6860.
- McCouch, S. R. *et al. Theor. appl. Genet.* (in the press).
- Bonierbale, M. W., Plaisted, R. L. & Tanksley, S. D. *Genetics* (in the press).
- Franklin, I. A. *Theor. Populat. Biol.*, **11**, 60-80 (1977).
- Stam, P. *Genet. Res.*, **25**, 131-155 (1980).
- Kosambi, D. D. *Ann. Eugen.*, **12**, 172-175 (1944).
- Tanksley, S. D. & Rick, C. M. *Theor. appl. Genet.*, **57**, 161-170 (1980).

Novel anti-inflammatory peptides from the region of highest similarity between uteroglobin and lipocortin I

Lucio Miele, Eleonora Cordella-Miele,
Antonio Facchiano & Anil B. Mukherjee*

Section on Developmental Genetics, Human Genetics Branch,
National Institute of Child Health and Human Development,
National Institutes of Health, Bethesda, Maryland, 20892, USA

Significant future developments in the effective treatment of inflammatory diseases may arise from non-toxic dual inhibitors of both cyclooxygenase and lipoxygenase pathways in the arachidonate cascade¹. Inhibition of phospholipase A₂ (PLA₂) (EC 3.1.1.4), may provide such a dual action and recent research has concentrated on the role of PLA₂-inhibitory proteins as possible anti-inflammatory agents. Blastokinin² or uteroglobin³ is a steroid-induced rabbit secretory protein with PLA₂-inhibitory activity. Its

biochemical and biological properties have been extensively studied⁴⁻¹⁴ and its crystallographic structure has been resolved at 1.34 Å (refs 15, 16). Lipocortins are a family of related proteins¹⁷⁻²², which, it has been suggested, mediate the anti-inflammatory effects of glucocorticoids (for a review, see ref. 23). Some proteins of this group have been purified^{24,25} and the complementary DNA sequences of two human lipocortins are known^{25,26}. Lipocortins inhibit PLA₂ *in vitro*^{17-21,24-29}, although their mechanism of action is still unclear^{24-29,30,31}. Recombinant lipocortin I inhibits eicosanoid synthesis in isolated perfused lungs from the guinea pig³². Here, we report that synthetic oligopeptides corresponding to a region of high amino-acid sequence similarity between uteroglobin and lipocortin I have potent PLA₂ inhibitory activity *in vitro* and striking anti-inflammatory effects *in vivo*.

Mature uteroglobin (UG) is an antiparallel dimer formed by two identical subunits of 70 amino acids each^{15,16}. Figure 1a shows the alignment of the mature UG monomer³³ with lipocortins I and II (refs 25, 26). Lipocortins I and II are formed by four non-identical repeated units of about 70 amino acids each³⁴⁻³⁶. The program PRTALN aligns UG with a region of lipocortins I and II that approximately corresponds to the second repeat (Fig. 1a, I and III), the region of highest similarity between the two lipocortins. The total similarity (identities + conservative replacements) is 40% in both cases. However, the highest number of identities (19) is found between UG and lipocortin I repeat three (residues 199-275), with a total similarity of 43%. Moreover, a striking local similarity was identified between residues 40-46 of UG and residues 247-253 of lipocortin I, repeat three. In UG, this region corresponds to the C-terminal part of α -helix three (residues 32-47)^{15,16}. As it is 'centered' on a relatively long region of local similarity, the alignment with repeat three is probably the most accurate. We drew four conclusions from these results. (1) There is amino-acid sequence similarity between UG and lipocortins; (2) UG is more similar to lipocortin I than to lipocortin II, although it can be aligned to the region of highest similarity between the two lipocortins; (3) compared to repeats two and three of lipocortin I, the highest density of identities and conservative replacements is in the C-terminal half of UG (residues 33-70) and particularly in α -helix three (Fig. 1a); (4) the region of most similarity between helix three of UG and repeat three of lipocortin I can be precisely identified as a heptapeptide spanning residues 40-46. This heptapeptide aligns with lipocortin I residues 247-253 (Fig. 1a).

Hydropathy profiles of UG and the corresponding regions of lipocortins are shown in Fig. 1b (I-IV). The similarity is evident over most of the UG sequence and particularly in the region of UG between positions 37-52. Interestingly, the hydropathy profile of porcine pancreatic PLA₂ (Fig. 1b, V) shows a striking similarity with that of the UG monomer and of the corresponding regions of lipocortins, even in the absence of significant sequence similarity. It has been independently suggested that lipocortin repeats and PLA₂ could have a similar three-dimensional organization³⁵. Moreover, refined crystallographic data have shown that the molecular surface of UG is strikingly similar to the one of PLA₂ (ref. 16). These data may indicate that PLA₂, the UG monomer and the lipocortin repeats have a similar three-dimensional organization.

On the basis of computer analyses, we synthesized oligopeptides corresponding to the C-terminal part of UG α -helix three and tested them for PLA₂-inhibitory activity. Table 1 shows the amino-acid sequences and PLA₂-inhibitory properties of synthetic peptides derived from UG and lipocortin I. Peptide one (P1), which corresponds to the nine C-terminal amino-acid residues of α -helix three, is a very potent inhibitor of PLA₂, with ~80% inhibition at 50 nM, under the experimental conditions used. Peptide two (P2), corresponding to lipocortin I residues 246-254, is as active as P1 (Table 2). Peptide three (P3), which retains full inhibitory activity under these experimental conditions, was constructed by substituting an

* To whom correspondence should be addressed.

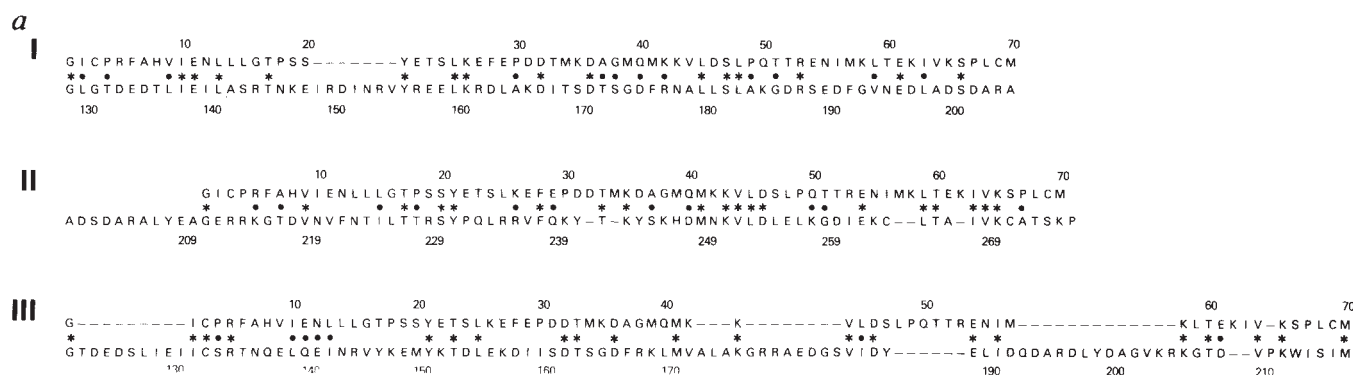
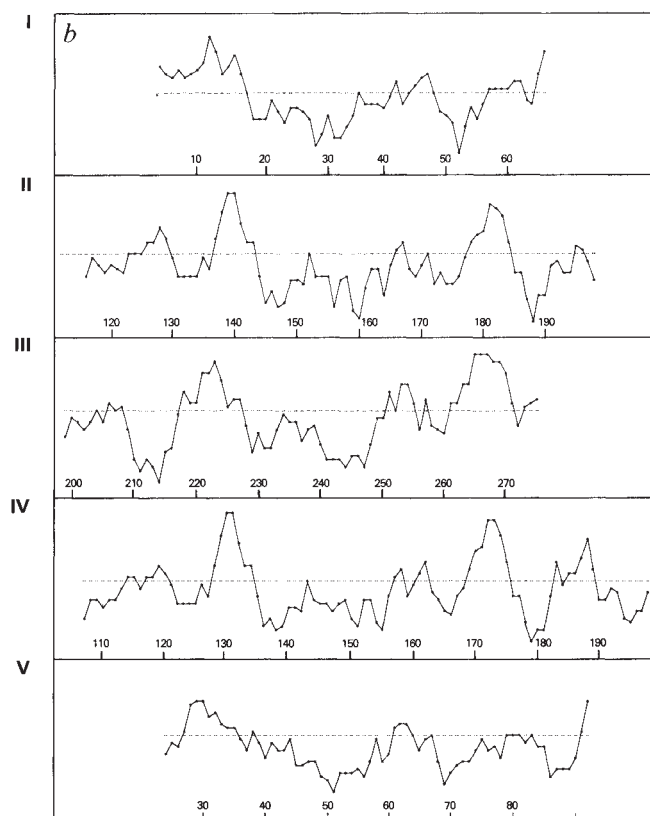


Fig. 1 *a*, Alignment of UG with lipocortin I repeat two (I), lipocortin I repeat three (II) and lipocortin II (III), obtained using the program PRTALN⁴⁴. Identities are indicated by asterisks and conservative substitutions by dots. Conservative substitutions were added manually and are defined by pairs of residues falling into the following groups⁴⁵. S, T, A, G, P; N, D, E, Q; R, K, H; M, I, L, V; F, Y, W. The sequence of mature UG is derived from ref. 33, the sequences of lipocortins I and II from refs 26 and 25 respectively. *b*, Hydropathy profiles of UG (I), lipocortin I repeat two (II) lipocortin I repeat three (III), lipocortin II repeat two (IV) and porcine pancreatic PLA₂, residues 24–92 (V). The profiles were obtained using a segment of seven residues. Positive peaks indicate hydrophobic regions; negative peaks indicate hydrophilic regions.

Methods. We used the following programs: PRTALN (ref. 44), FASTP (ref. 46), RDF (Ref. 46) and HYDRO. HYDRO was written by Dr C. Buckler for the NIH Molecular Biology Users Group. The program generates hydropathy profiles⁴⁷. Similarity scores and statistical evaluations were obtained by means of FASTP and RDF. Briefly, FASTP calculates a similarity score before and after optimization of the alignment. The algorithm is similar to the one used in PRTALN, but the scoring rules are based on the 250-PAM amino acid replaceability matrix^{45,46}. 'Rare' amino-acid matches get higher scores than 'common' amino-acid matches. Conservative substitutions also get positive scores, whereas 'unlikely' substitutions and mismatches get a negative score. RDF evaluates the significance of a similarity between two sequences by comparing the first sequence with n (Usually 100) random shuffles of the second. Initial and optimized scores are calculated for every comparison according to the scoring rules of FASTP. The significance is expressed as a z -value ($z = (s - m)/sd$, where s is the score of the original comparison, m is the mean score of the random-shuffled controls and sd is the standard deviation of this mean). z values between three and six are considered to be possibly significant for an evolutionary relationship, between 6 and 10 probably significant and above 10 certainly significant⁴⁶. As a control, we used FASTP to search the 3061 sequences of the Protein Identification Resource database⁴⁸ for similarity with UG. The 111 alignments with highest scores were identified and saved. The average optimized score of these alignments is 36.09 ± 6.8 . Lipocortins I and II received scores of 42 and 39, which are respectively 0.9 and 0.4 standard deviations higher than this average. We also used FASTP and RDF to calculate the similarity scores and z values obtained comparing UG to all other proteins that have been suggested to be evolutionarily related to it. These include rat prostatic steroid binding protein chain C2 and Ca²⁺ binding proteins of the calmodulin family^{49,50}. Lipocortins I and II have optimized z values which are higher than all other proteins tested except prostatic steroid binding protein C2 chain precursor (not shown). When UG α -helix three alone was compared to lipocortin I repeat three initial and optimized z values were 4.24 and 3.03 standard deviations respectively.



asparagine for a lysine residue corresponding to UG lysine 42. The removal of two amino-acid residues from the N-terminal and one residue from the C-terminal of P1 and P2 abolishes the inhibitory activity (Table 1). The 'core' tetrapeptide Lys-Val-Leu-Asp, which is common to all the active peptides, is itself inactive. Peptide seven, corresponding to UG residues 1–10 (a region of lesser similarity between UG and lipocortins) has been used as a non-specific control and is inactive as a PLA₂ inhibitor. No significant difference in PLA₂ inhibitory activity between peptides one, two and three could be detected under these experimental conditions, at peptide concentrations between 1 and 200 nM (not shown).

We further investigated the PLA₂-inhibitory properties of the peptides and their parent proteins under initial velocity conditions. Figure 2 shows that UG, P1 and P2 have comparable dose

responses. However, UG had a maximal inhibitory effect at 1 nM, whereas the peptides (P1 and P2) had a similar effect at concentrations between 4 and 8 nM. Under these conditions, both UG and the peptides had a remarkable inhibitory effect in sub-nanomolar concentrations (Fig. 2). Decreasing the enzyme concentration caused a parallel decrease in the optimum inhibitory concentrations of peptides. When P1 was pre-incubated with the lipid substrate instead of the enzyme, its inhibitory effect was drastically reduced (Fig. 2a). This may indicate that the inhibitory effect of P1 is exerted through an interaction with the enzyme rather than the substrate. This possibility is further supported by the observation that P1, at concentrations ranging from 1 nM to 10 μ M, had no inhibitory effect on *Bacillus cereus* phospholipase C, under identical experimental conditions to those used for PLA₂ (data not shown). Figure 2b clearly shows

Table 1 Amino-acid sequences of synthetic peptides and their PLA₂ inhibitory activity

Peptide	Sequence	Concentration* (M)	% PLA ₂ inhibition† ± standard deviation
1	MQMKKVLDS	5×10^{-8}	81.6 ± 5.3
2	HDMNKVLDL	5×10^{-8}	89.6 ± 1.1
3	MQMNKVLDS	5×10^{-8}	86.5 ± 0.8
4	KVLD	10^{-3}	n.s.
5	MKKVLD	10^{-3}	n.s.
6	MNKVLD	10^{-3}	n.s.
7	GICPRFAHVI	10^{-3}	n.s.

PLA₂ was assayed according to Clark *et al.*⁴⁰ with minor modifications. Briefly, the reaction mixture contained 100 mM Tris HCl, pH 8.0, 100 mM NaCl, 1 mM CaCl₂, 1 mM sodium deoxycholate, 10 μ M 1-stearoyl, 2-[1-¹⁴C] arachidonyl phosphatidylcholine (58 mCi mmol⁻¹; Amersham) and 25 ng (approximately 36 nM) porcine pancreatic PLA₂ (Boehringer, 700 U mg⁻¹) in a total volume of 50 μ l. The reaction was started by addition of the enzyme to the radioactive substrate. Reactions were run at 37 °C for 4 min. Approximately 20–30% of the substrate was hydrolysed. Appropriate controls were included with one without enzyme. To determine the inhibitory activity, PLA₂ was pre-incubated for 10 min with the putative inhibitors at 37 °C before adding the enzyme/inhibitor mixture to the substrate. There were also controls in which PLA₂ was pre-incubated with buffer. Radioactive arachidonic acid was separated from the substrate by TLC on silica plates (silica gel G, Analtech). Developing solvent was petroleum ether/ethyl ether/acetic acid (70:30:1). Iodine-stained bands co-migrating with authentic arachidonic acid were scraped and counted in a Beckman LS-9000 liquid scintillation counter. *, The concentration used was that yielding highest inhibition for active peptides; the highest concentration tested was used for inactive peptides. †, Values are means of three determinations. n.s., not significant.

that P3 under these conditions is less active than P1 and P2. This difference tends to decrease when the concentration of P3 is increased. Recombinant lipocortin I inhibits PLA₂ in this system, with an overall dose-response curve similar to that of P3. Also, lipocortin I seems to be less active than the lipocortin I-derived peptide (P2). However, this difference may be artefactual, as it has been reported that purified preparations of recombinant lipocortin I are mixtures of several disulphide-bonded forms and are contaminated with bacterial proteins³².

Several conclusions can be drawn from these experiments. The first two residues of P1, and the last one, can be replaced by other amino acids, but not eliminated, without loss of activity. This may indicate that the length of the peptide is critical, possibly for conformational reasons; the minimum peptide length that can form two complete turns of an α -helix is eight residues. The lysine residue in P1 corresponding to lysine 42 of UG is not indispensable for the inhibitory activity. This was expected, because in UG, lysine 42 is not exposed to the solvent, as it is H-bonded to the main-chain carbonyl of glycine 16 (ref. 15). However, hybrid peptide three seems to be less active than both parent peptides one and two, particularly at low concentrations, possibly because there is a decreased tendency to assume the active conformation in solution. Recombinant lipocortin I inhibits PLA₂ in our system, although with a lower activity than UG or P1, in the absence of phosphatidylserine and in the same concentration range as the enzyme. Both the presence of phosphatidylserine and a large molar excess of lipocortin over PLA₂ have been reported to be necessary for the inhibition of PLA₂ by lipocortin in assays using autoclaved *Escherichia coli* cells or extracted *E. coli* phospholipids as enzyme substrates^{30,31}. The mechanism of PLA₂ inhibition by lipocortin I in our assay system is currently under investigation. The similarity in the PLA₂-inhibitory properties of P1 and purified UG support the hypothesis that P1 corresponds to an active site, or part of an active site, responsible for the PLA₂-inhibitory activity of UG.

We tested the PLA₂-inhibitory synthetic peptides for anti-inflammatory activity *in vivo* and found that both UG-derived P1 and lipocortin I-derived P2 have very potent anti-inflamma-

Table 2 Anti-inflammatory effect of peptides 1 and 2 on carrageenan-induced rat paw oedema

Treatment	Number of animals	% Inhibition (mean ± standard deviation)
P1, 2 mg kg ⁻¹	7	82.6 ± 5 (<i>P</i> < 0.01)
P1, 200 μ g kg ⁻¹	9	57.5 ± 7 (<i>P</i> < 0.01)
P1, 20 μ g kg ⁻¹	9	36.0 ± 8 (<i>P</i> < 0.01)
P1, 2 μ g kg ⁻¹	9	31.5 ± 9 (<i>P</i> < 0.01)
P1, 0.2 μ g kg ⁻¹	9	24.0 ± 5 (<i>P</i> < 0.01)
P1, 0.02 μ g kg ⁻¹	9	n.s. (<i>P</i> > 0.05)
P2, 2 mg kg ⁻¹	8	96.2 ± 2 (<i>P</i> < 0.01)
P2, 200 μ g kg ⁻¹	8	30.5 ± 7 (<i>P</i> < 0.01)
P2, 20 μ g kg ⁻¹	8	36.3 ± 7 (<i>P</i> < 0.01)
P2, 2 μ g kg ⁻¹	8	23.3 ± 3 (<i>P</i> < 0.01)
P2, 0.2 μ g kg ⁻¹	8	20.0 ± 4 (<i>P</i> < 0.01)
P2, 0.02 μ g kg ⁻¹	8	12.2 ± 3 (<i>P</i> < 0.05)
DEX, 1 μ g kg ⁻¹	5	75.9 ± 6 (<i>P</i> < 0.01)
IND, 1 mg kg ⁻¹	5	26.1 ± 4 (<i>P</i> < 0.01)
P1, 2 mg kg ⁻¹ + AA	4	21.1 ± 5 (<i>P</i> < 0.01)
P2, 2 mg kg ⁻¹ + AA	4	20.2 ± 1 (<i>P</i> < 0.01)
UG, 100 μ g kg ⁻¹	4	35.5 ± 7 (<i>P</i> < 0.01)
LC-1, 100 μ g kg ⁻¹	4	24.5 ± 1 (<i>P</i> < 0.01)
P7, 2 mg kg ⁻¹	4	n.s. (<i>P</i> > 0.05)
P7, 10 μ g kg ⁻¹	4	n.s. (<i>P</i> > 0.05)
BSA, 2 mg kg ⁻¹	4	n.s. (<i>P</i> > 0.05)
BSA, 100 μ g kg ⁻¹	4	n.s. (<i>P</i> > 0.05)
LSZ, 2 mg kg ⁻¹	4	n.s. (<i>P</i> > 0.05)
LSZ, 100 μ g kg ⁻¹	4	n.s. (<i>P</i> > 0.05)

P1, peptide 1; P2, peptide 2; DEX, dexamethasone; IND, indometacin; AA, arachidonic acid; UG, uteroglobin; LC-1, lipocortin I; P7, peptide 7; BSA, bovine serum albumin; LSZ chicken egg lysozyme; n.s., not significant. Rats (Sprague-Dawley, males, mass about 250 g) were maintained with water and food *ad libitum*, with a 12-hour-light/12-hour-darkness cycle. Animals were injected in the sub-plantar space with 1.0 mg lambda-carrageenan (Sigma, type IV) in 0.1 ml sterile saline. Inhibitors, control substances or vehicle alone were injected about 30 s after carrageenan, in a volume of 0.1 ml, to avoid non-specific interactions *in vitro* between carrageenan and inhibitors. Controls which received saline alone and saline plus vehicle were included. Peptides were dissolved in sterile 10 mM Tris, pH 8, because this yielded more consistent and reproducible dose responses compared to peptides dissolved in saline. Dorsoplantar paw thickness was measured with a vernier caliper^{41,42} immediately before the carrageenan injection and 4 h after treatment. When AA was used, it was administered 10 μ g per paw. At this dosage, AA did not cause appreciable paw swelling when administered alone, nor did it increase the swelling caused by carrageenan when administered with it (*N*, four animals per group). Inhibitory effects were assessed by comparing the dorsoplantar paw thickness of inhibitor-treated groups to that of vehicle-treated groups. The results were analysed by a one-tailed Student's *t* test for groups of unpaired observations. Significance was taken at *P* < 0.05. The statistical significance of the effects of P1 and P2 was also confirmed by one-way ANOVA⁴³. Note that the dose-response curve of P1 is steeper than that of P2 and that at the highest dose tested P2 seems to have a more pronounced effect than P1.

tory effect on the carrageenan-induced rat footpad oedema³⁷. Table 2 shows the anti-inflammatory effects of P1 and P2 compared to those of known anti-inflammatory agents. Both P1 and P2, when injected locally immediately after carrageenan administration, inhibited the formation of inflammatory oedema over a wide range of doses. The dose-response curves of P1 and P2 do not differ significantly, except perhaps in the range between 0.2 and 2 mg kg⁻¹. At 2 mg kg⁻¹ both peptides caused a virtually complete suppression of the inflammatory response. Both parent proteins, UG and lipocortin I, when used at 100 μ g kg⁻¹ had anti-inflammatory activity comparable to that of a 10-fold higher dose of indometacin. The effect of both P1 and P2, at the highest dose used, is drastically reduced by a concomitant local administration of 10 μ g of arachidonic acid (Table 2). This dosage of arachidonic acid had no inflammatory effect when administered alone and did not significantly increase the inflammatory effect of carrageenan when administered with

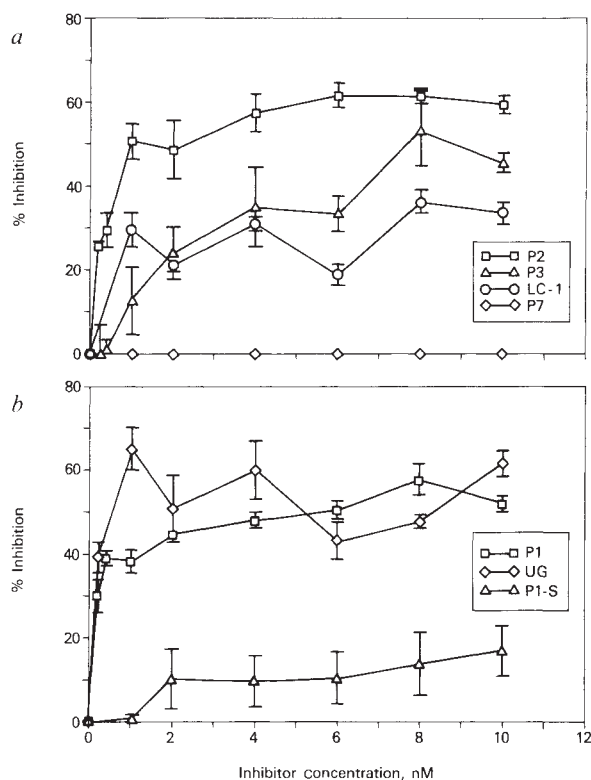


Fig. 2 Inhibition of porcine pancreatic PLA₂ by UG, lipocortin I and synthetic oligopeptides. Each point represents the mean of at least four separate determinations $\pm$ standard deviation. *a*, $\square$, peptide 2; $\triangle$, peptide 3; $\circ$, lipocortin-I; $\diamond$, peptide 7. *b*, $\square$, Peptide 1; $\diamond$, uteroglobin; $\triangle$, peptide 1, pre-incubated with the lipid substrate.

Methods. Peptides were synthesized by an Applied Biosystems model 430 A peptide synthesizer. Coupling efficiency was determined by ninhydrin monitoring. Cleavage of peptides from the solid support and removal of the side-chain protecting groups were performed by an optimized hydrofluoric acid method. Peptides were purified by reverse-phase HPLC, on a C8 stationary phase. Purity was determined by HPLC and amino-acid analysis. Lyophilized peptides were stored at -20°C under nitrogen in sealed glass vials and were redissolved immediately before use. UG was purified essentially as described⁸ from the uterine flushings of rabbits pre-treated with human chorionic gonadotropin. The protein seemed homogeneous as judged by SDS-polyacrylamide gel electrophoresis and isoelectric focusing (pI 5.4) and was stored desiccated at -20°C in lyophilized form. PLA₂ assays were performed as described in Table 1, with 2 nM PLA₂ and a reaction time of 30 s. Under these conditions, enzyme activation is virtually instantaneous (<10 s), and is followed by an interval during which the rate of product accumulation is linear (L. Miele *et al.* manuscript in preparation). The concentration of enzyme used falls within the linear part of the velocity against enzyme concentration curve. In 30 s 0.5–1% of the substrate is hydrolysed. This allows the effects on PLA₂ activity of the accumulation of reaction product in the lipid-water interface to be minimized. Non-specific proteins (horse heart myoglobin and chicken egg lysozyme) were tested for PLA₂-inhibitory activity at concentrations between 1 nM and 10 μM and were found to be inactive. In some experiments, phospholipase C from *Bacillus cereus* (0.01 U, Boehringer, grade I) was substituted for PLA₂. In these experiments the reaction time was 4 min. TLC separation of the reaction products was carried out in the same developing system used for PLA₂, and bands co-migrating with the standard (1,2-dioleoylglycerol) were scraped and counted in a Beckman LS-9000 liquid-scintillation counter.

it. These observations support the hypothesis that an important *in vivo* mechanism of these peptides may be the inhibition of arachidonate release from the cell membrane by PLA₂. However, our data do not rule out a minor contribution from other unknown mechanism(s) at the highest dosages of P1 and P2. Preliminary data indicate that P1 and P2 also have potent anti-inflammatory effects when administered by intraperitoneal and intravenous injection.

The evolutionary origin of PLA₂ inhibitory proteins is unknown. It has been suggested recently that lipocortins and calelectrins are derived from a 'single repeat' protein by gene duplication³⁸. This protein might also be the ancestor of the UG monomer. However, we cannot determine from our data whether the similarities arose from convergent or divergent evolution. We believe that the peptides described in this paper may be useful in future studies on the mechanism of action of PLA₂ inhibitory proteins. All we can say at present is that the much higher PLA₂ inhibitory activity observed when P1 is pre-incubated with PLA₂ than with the lipid substrate, the apparent specificity of P1 for PLA₂, compared with phospholipase C and the very low inhibitor:substrate molar ratios at which PLA₂ inhibition is observed in our system indicate, at least for P1, a different mechanism of action from the 'lipid coating' model^{30,31}. The construction of PLA₂-inhibitory oligopeptides from UG α -helix three and lipocortin I repeat three, on the basis of their high local similarity may indicate that similar sites of these proteins are responsible for their PLA₂ inhibitory activity. However, the region identified by our approach is not the only possible active site of lipocortins. We have not investigated whether either repeat two of lipocortin I, which may be necessary for the inhibitory activity²⁹, or lipocortin II, contain other PLA₂-inhibitory regions. However, when lipocortin I repeat three was aligned to lipocortin I repeat two and to lipocortin II repeat two by PRTALN, the region between residues 246–254 was aligned to positions 162–170 of lipocortin I and to positions 153–161 of lipocortin II.

Release and metabolism of arachidonate by mast cells are among the earliest biochemical changes in the inflammatory

response to carrageenan³⁹. They occur in response to carrageenan-induced 'cytoplasmic injury' of these cells³⁹. Additionally, it has been suggested that liberation of PLA₂ from neutrophils participates in the amplification of the inflammatory response to carrageenan³⁹. The *in vivo* pharmacological effects of PLA₂-inhibitory peptides make them a valuable model for the development of novel anti-inflammatory agents of therapeutic importance. Because the anti-inflammatory effect is the first of their biological properties to be identified, we propose the name 'antiflammins' for the peptides described here.

We thank Dr Jim Sidbury Jr for his support and suggestions Drs H. C. Chen and J. DeB. Butler for discussions and for critical review of the manuscript, and Dr Blake Pepinsky for the gift of recombinant lipocortin I. We wish to acknowledge the technical support provided by Dr Anita Hong and Mr Michael Brasseur (Applied Biosystems) and Dr Martha Knight (Peptide Biotechnologies, Inc.), who prepared the synthetic peptides used in this study.

Received 8 July; accepted 9 September 1988.

1. Vane, J. R. & Botting, R. *FASEB J.* **1**, 89–96 (1987).
2. Krishnan, R. S. & Daniel, J. C. *Jr Science* **158**, 490–492 (1967).
3. Beier, H. M. *Biochim. biophys. Acta* **160**, 289–291 (1968).
4. Beato, M. *et al.* *Regulation of Gene Expression by Hormones* (ed. McKerns, K. W.) 151–175 (Plenum, New York, 1983).
5. Miele, L., Cordella-Miele, C. & Mukherjee, A. B. *Endocr. Rev.* **8**, 474–490 (1987).
6. Mukherjee, A. B., Cordella-Miele, E., Kikukawa, T. & Miele, L. *Modification of Proteins and Ageing, Advances in Experimental Medicine and Biology* (eds Zappia, V., Galletti, P., Porta, R. & Wold, F.) **231**, 135–151 (1988).
7. Mukherjee, D. C., Ulane, R. E., Manjunath, R. & Mukherjee, A. B. *Science* **219**, 989–991 (1983).
8. Mukherjee, A. B., Ulane, R. E., & Agrawal, A. K. *Am. J. reprod. Immunol.* **2**, 135–141 (1982).
9. Mukherjee, A. B., Cunningham, D. C., Agrawal, A. K. & Manjunath, R. *Ann. N.Y. Acad. Sci.* **392**, 401–402 (1982).
10. Vasanthakumar, G., Manjunath, R., Mukherjee, A. B., Warabi, H. & Schiffman, E. *Biochem. Pharmacol.* **37**, 389–394 (1988).
11. Schiffman, E. *et al.* in *Agents and Action Supplements Vol. 2* (eds Keller, H. & Till, E.D.) 106–120 (Birkhauser, Basel, 1983).
12. Schiffman, E. *et al.* in *Asthma: Physiology, Immunopharmacology and Treatment. Third International Symposium* (eds Kay, A. B., Austen, K. F. & Lichtenstein, L. M.) 173–193 (Academic, London, 1984).

13. Manjunath, R. *et al.* *Biochem. Pharmacol.* **36**, 741-746 (1987).
14. Levin, S. W., Butler, J. D., Schumacher, U. K. & Mukherjee, A. B. *Life Sci.* **38**, 1813-1819 (1986).
15. Mornon, J. P., Fridlansky, F., Bally, R. & Milgrom, E. *J. molec. Biol.* **137**, 415-429 (1980).
16. Morize, I. *et al.* *J. molec. Biol.* **194**, 725-739 (1987).
17. Flower, R. J. & Blackwell, G. J. *Nature* **278**, 456-459 (1979).
18. Blackwell, G. J. *et al.* *Nature* **287**, 147-149 (1980).
19. Hirata, F., Schiffman, E., Venkatasubramanian, K., Salomon, D. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2533-2536 (1980).
20. Russo-Marie, F. & Duval, D. *Biochim. biophys. Acta* **712**, 177-185 (1982).
21. Cloix, J. F., Colard, O., Rothut, B. & Russo-Marie, F. *Br. J. Pharmac.* **79**, 313-321 (1983).
22. Di Rosa, M., Flower, R. J., Hirata, F., Parente, L. & Russo-Marie, F. *Prostaglandins* **28**, 441-442 (1984).
23. Flower, R. J. *Adv. Inflamm. Res.* **8**, 1-34 (1984).
24. Pepinsky, R. B. *et al.* *J. biol. Chem.* **261**, 4239-4246 (1986).
25. Huang, K. S. *et al.* *Cell* **46**, 191-199 (1986).
26. Wallner, B. P. *et al.* *Nature* **320**, 77-81 (1986).
27. Hirata, F. *J. biol. Chem.* **256**, 7730-7733 (1981).
28. Hirata, F. in *Advances in Prostaglandin, Thromboxane and Leukotriene Research* Vol. 11 (eds Samuelsson, B., Paoletti, R. & Ramwell, P.) 73-78 (Raven, New York 1983).
29. Huang, K. S. *et al.* *J. biol. Chem.* **262**, 7639-7645 (1987).
30. Davidson, F. F., Dennis, E. A., Powell, M. & Glenney, J. R. Jr *J. biol. Chem.* **262**, 1698-1705 (1987).
31. Haigler, H. T., Schlaepfer, D. D. & Burgess, W. H. *J. biol. Chem.* **262**, 6921-6930 (1987).
32. Cirino, G., Flower, R. J., Browning, J. L., Sinclair, L. K. & Pepinsky, R. B. *Nature* **328**, 270-272 (1987).
33. Chandra, T., Bullock, D. W. & Woo, S. L. C. *DNA* **1**, 19-25 (1981).
34. Weber, K. & Johnson, N. *FEBS Lett.* **203**, 95-98 (1986).
35. Geisow, M. J. *FEBS Lett.* **203**, 99-103 (1986).
36. Munn, T. Z., Mues, G. I. *Nature* **322**, 314-315 (1986).
37. Winter, C. A., Risley, E. A. & Nuss, G. W. *J. pharmac. exp. Ther.* **141**, 369-376 (1963).
38. Sudhoff, T. C., Slaughter, C. A., Leznicki, I., Barion, P. & Reynolds, G. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 664-668 (1988).
39. Vinegar, R. *et al.* *Fedn Proc.* **46**, 118-126 (1987).
40. Clark, M. A. *et al.* *J. biol. Chem.* **261**, 10713-10718 (1986).
41. Newbould, B. B. *Br. J. Pharmac.* **21**, 127-136 (1963).
42. Hargreaves, K., Dubner, R., Brown, F., Flores, C. & Joris, J. *Pain* **32**, 77-88 (1988).
43. Snedecor, G. W. & Cochran, W. G. in *Statistical Methods* 7th edn (Iowa State University Press, Ames, 1980).
44. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726-730 (1983).
45. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. in *Atlas of Protein Sequence and Structure* Vol. 5, suppl. 3 (ed. Dayhoff, M. O.) 345-352 (National Biomedical Research Foundation, Washington, DC 1978).
46. Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435-1441 (1985).
47. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
48. Barker, W. C. *et al.* *Protein Identification Resource Data Base* (National Biomedical Research Foundation, Washington, DC 1985).
49. Baker, M. E. *Biochem. biophys. Res. Commun.* **114**, 325-330 (1983).
50. Baker, M. E. *FEBS Lett.* **189**, 188-194 (1985).

Table 1 Strains of mice and their H-2 haplotypes

Strain	Abbreviation	H-2 subregions						
		K	A α	A β	E β	E α	S	D
C57B1/6J	B6	b	b	b	b	b	b	b
		b	b	b	b	b	b	b
(C57B1/6J × DBA/2) F ₁	(B6 × DBA/2) F ₁	b	b	b	b	b	b	b
		d	d	d	d	d	d	d
B10.HTG	HTG	d	d	d	d	d	d	b
		d	d	d	d	d	d	b
B10.D2 (R107)	R107	b	b	b	b	b	b	d
		b	b	b	b	b	b	d
B10.BR	B10.BR	k	k	k	k	k	k	k
		k	k	k	k	k	k	k

In the thymus of H-2^b $\alpha\beta$ TCR transgenic female mice expressing a TCR for the male (H-Y) antigen in the context of class I (H-2D^b) MHC molecules, the proportion of single positive CD4⁺8⁺ thymocytes is significantly elevated as compared with normal nontransgenic littermates^{5,6}. The increased proportion of CD4⁺8⁺ thymocytes resulted from positive selection caused by the interaction of CD4⁺8⁺ thymocytes expressing the transgenic TCR with MHC molecules encoded by the H-2^b but not the H-2^d haplotype⁵. This was shown by back-crossing the C57B1/6 × DBA/2 F₂ transgenic founder mouse to either C57L or DBA/2 mice and by analysing CD4⁺8⁺ thymocytes expressing the $\alpha\beta$ transgenic receptor in H-2^{b/b} and H-2^{d/d} homozygous and in H-2^{d/b} heterozygous offspring. To investigate further the specificity of the positive selection exerted by MHC molecules, we performed thymus repopulation experiments: lethally irradiated females of the intra-H-2 recombinant strains B10.HTG (HTG) and B10.D2 (R107), which differ at D and K through to S subregions of the H-2 complex, were transplanted with bone marrow cells from H-2^b $\alpha\beta$ transgenic female mice. Control groups included syngeneic, H-2 allogeneic and semi-allogeneic (F₁) recipients of bone marrow cells from normal, β transgenic and $\alpha\beta$ transgenic mice. The recipient strains and their H-2 haplotypes are listed in Table 1. Five weeks after transplantation, the composition of CD4/CD8 thymocyte subpopulations in repopulated thymuses and the expression level of transgenic TCR proteins were analysed by staining with anti-CD4, anti-CD8, F23.1 (ref. 7), and with the recently obtained monoclonal T3.70 antibody. The T3.70 antibody is specific for an idiotype determinant dependent on the expression of both α and β transgenic TCR chains (ref. 5 and unpublished results; see also Fig. 1).

The repopulating capacity of donor stem cells, as estimated by the cellularity of the thymuses of the recipients, correlated with the degree of compatibility within the H-2 complex: the highest numbers of thymocytes ($\sim 50 \times 10^6$ per thymus) were recovered from syngeneic and semi-allogeneic recipients, the lowest ($\sim 8 \times 10^6$ per thymus) from allogeneic recipients differing at the whole H-2 complex (Table 2). In general, $\alpha\beta$ transgenic bone marrow stem cells yielded one third to one half of the progeny compared with bone marrow from normal mice. As shown in the last row in Fig. 1, more than 98% of thymocytes in the recipients of transgenic bone marrow stained with F23.1 monoclonal antibody which is specific for the transgenic β chain of the TCR (ref. 6), indicating that almost all thymocytes were derived from the donor stem cells. In all groups of mice repopulated by non-transgenic stem cells and in non-D^b recipients (B10.BR and R107) of transgenic stem cells, as well as in B6 (H-2^b) mice repopulated by β transgenic stem cells, the relative proportions of CD4⁺8⁻ to CD4⁺8⁺ thymocytes were in the normal range (Fig. 1, two upper rows): the ratio of CD4⁺8⁻ to CD4⁺8⁺ thymocytes varied from 3.0 to 6.5 (Table 2). On the other hand, a significantly lower ratio of CD4⁺8⁻ to CD4⁺8⁺

Positive selection of antigen-specific T cells in thymus by restricting MHC molecules

Pawel Kisielow*†, Hung Sia Teh*†, Horst Blüthmann‡ & Harald von Boehmer*§

* Basel Institute for Immunology, 487 Grenzacherstrasse, CH-4058 Basel, Switzerland

‡ Central Research Units, F. Hoffmann-La Roche & Co. Ltd, CH-4002 Basel, Switzerland

Thymus-derived lymphocytes (T cells) recognize antigen in the context of class I or class II molecules encoded by the major histocompatibility complex (MHC) by virtue of the heterodimeric $\alpha\beta$ T-cell receptor (TCR)^{1,2}. CD4 and CD8 molecules expressed on the surface of T cells bind to nonpolymorphic portions of class II and class I MHC molecules and assist the TCR in binding and possibly in signalling^{3,4}. The analysis of T-cell development in TCR transgenic mice has shown that the CD4/CD8 phenotype of T cells is determined by the interaction of the $\alpha\beta$ TCR expressed on immature CD4⁺8⁺ thymocytes with polymorphic domains of thymic MHC molecules in the absence of nominal antigen⁵. Here we provide direct evidence that positive selection of antigen-specific, class I MHC-restricted CD4⁺8⁺ T cells in the thymus requires the specific interaction of the $\alpha\beta$ TCR with the restricting class I MHC molecule.

† Permanent addresses: The Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wrocław, Poland (P.K.); Department of Microbiology, University of British Columbia, Vancouver V6T 1W5, Canada (H.S.T.).

§ To whom correspondence should be addressed.

Table 2 Number of thymocytes and proportions of CD4⁺8⁺ and CD4⁺8⁻ subpopulations in repopulated thymuses of bone marrow irradiation chimaeras

Female donor of bone marrow	Female recipient (850R)	Number of thymocytes $\times 10^6$	%CD4 ⁺ 8 ⁻	%CD4 ⁺ 8 ⁺	CD4 ⁺ 8 ⁻ /CD4 ⁺ 8 ⁺
B6	B6	48; 60 (54)*	12.9; 11.1 (12.0)	1.6; 2.8 (2.2)	7.9; 3.9 (5.9)
HTG	HTG	24; 46 (35)	10.1; 10.7 (10.4)	2.0; 2.1 (2.0)	5.0; 5.1 (5.0)
R107	R107	54; 60 (57)	8.5; 8.4 (8.4)	1.6; 1.8 (1.7)	5.3; 4.6 (4.9)
B6	B10.BR	24; 8 (16)	15.9; 28.2 (22.1)	7.4; 7.3 (7.3)	2.1; 3.9 (3.0)
B6	(B6 $\times$ DBA/2) F ₁	50; 50 (50)	11.4; 10.2 (10.8)	1.8; 1.5 (1.6)	6.2; 6.8 (6.5)
B6	HTG	4; 32 (18)	78.2; 8.5 (43.3)	12.5; 1.6 (7.0)	6.2; 5.3 (5.7)
B6	R107	31; 31 (31)	6.6; 6.6 (6.6)	1.7; 1.9 (1.8)	4.0; 3.5 (3.7)
$\alpha\beta$	B6	21; 13 (17)	10.4; 10.1 (10.2)	17.7; 13.9 (15.8)	0.6; 0.7 (0.6)
$\alpha\beta$	B10.BR	6; 8 (7)	31.3; 17.0 (24.1)	7.0; 3.4 (5.2)	4.4; 5.0 (4.7)
$\alpha\beta$	(B6 $\times$ DBA/2)F ₁	20; 12 (16)	12.2; 8.5 (10.3)	7.6; 4.9 (6.2)	1.6; 1.7 (1.6)
$\alpha\beta$	HTG	16; 16 (16)	8.3; 8.0 (8.1)	18.2; 16.7 (17.4)	0.4; 0.5 (0.4)
$\alpha\beta$	R107	8; 9 (8)	17.5; 12.7 (15.1)	6.2; 2.5 (4.3)	2.8; 5.0 (3.9)
β	B6	31; 20 (25)	14.0; 14.7 (14.3)	2.6; 2.5 (2.5)	5.4; 5.9 (5.6)

Irradiation chimaeras were prepared and analysed as described in Fig. 1. Two mice were analysed in each group.

* Mean value.

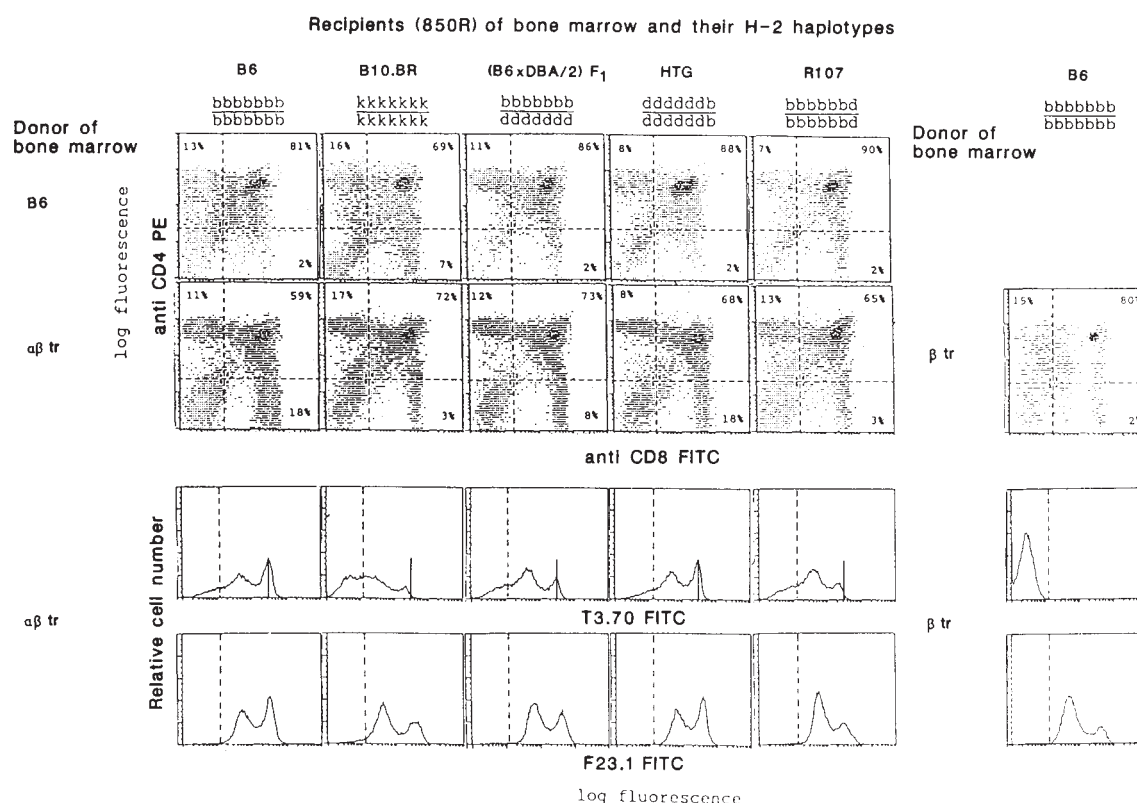


Fig. 1 Surface expression of CD4, CD8 and $\alpha\beta$ transgenic (T3.70⁺; F23.1⁺) or β transgenic (T3.70⁻; F23.1⁺) TCR on thymocytes from various bone marrow irradiation chimaeras. Thymocytes were stained with a mixture of phycoerythrin (PE)-conjugated anti-CD4 and fluorescein (FITC)-conjugated anti-CD8 monoclonal antibodies mAbs (two upper rows) or by unconjugated T3.70 or F23.1 mAbs followed by FITC-conjugated (Fab')₂ fragments of sheep anti-mouse immunoglobulin antibody (two lower rows). Vertical bars on histograms representing the staining with T3.70 mAb mark the population of cells expressing high levels of T3.70⁺ TCR, present in D^b-positive, but absent in D^b-negative, recipients of $\alpha\beta$ transgenic bone marrow cells.

Methods. Lethally irradiated (850R) females of the strains of mice indicated were injected i.v. with anti-Thy1 + C treated (37 °C, 45 min) bone marrow cells (5×10^6 per recipient) from B6, $\alpha\beta$ transgenic or β transgenic female mice. Five weeks later, mice were killed, whole thymuses carefully removed, single-cell suspensions prepared, counted and stained with phycoerythrin-conjugated anti-CD4 (Becton Dickinson), FITC-conjugated anti-CD8 (Becton Dickinson), T3.70 and F23.1 mAbs, as indicated above and described in detail elsewhere⁵. Two- or single-colour fluorescence analysis was performed on FACScan (Becton Dickinson) flow cytometer with logarithmic intensity scales using FACScan research software program (FRSP). Ten thousand viable cells were analysed in each sample. The results of two-colour stainings are presented as density plots generated by analysis of processed data reduced to a 64×64 matrix with 16 levels. Percentages of single- or double-stained cells were calculated using FRSP. Markers were set against control samples which included substitution of diluent or second-step reagent for either one or both antibodies.

Recipients (850R) of bone marrow
from $\alpha\beta$ transgenic female
and their H-2 haplotypes

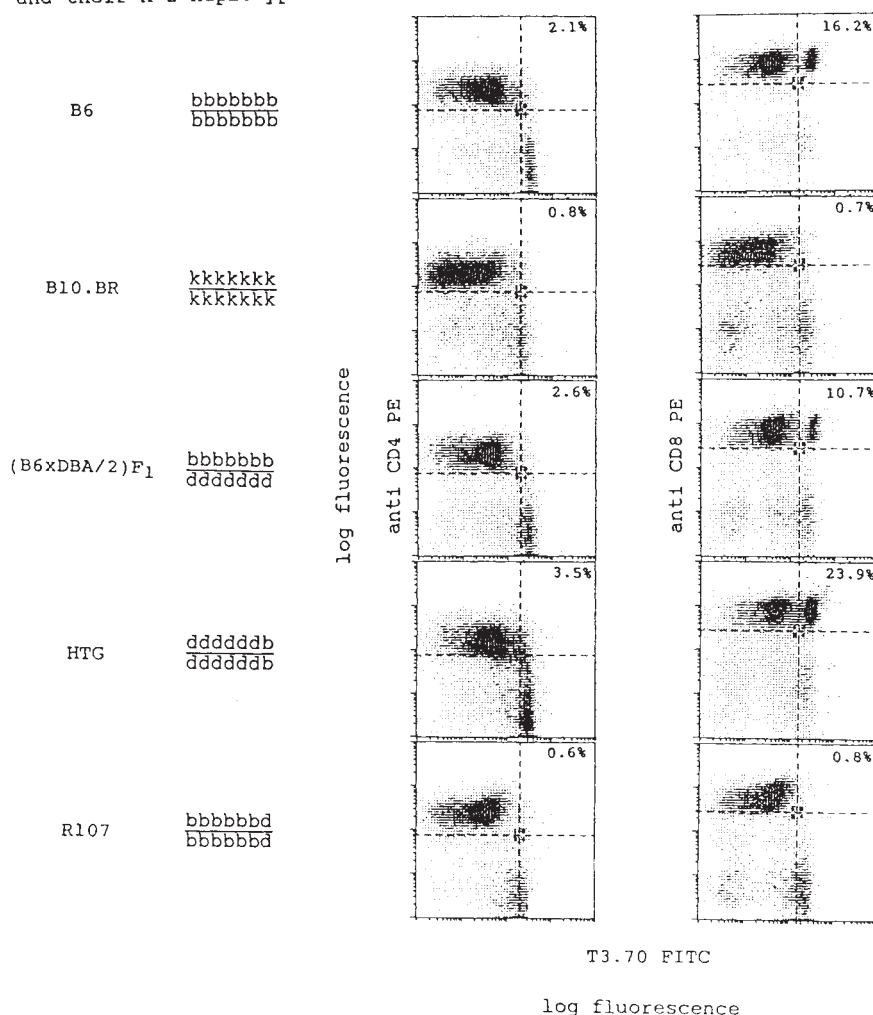


Fig. 2 Expression of $\alpha\beta$ transgenic (T3.70⁺) TCR on CD4/CD8 subpopulations of thymocytes from various irradiation chimaeras reconstituted with bone marrow of $\alpha\beta$ TCR transgenic females. Thymocytes were incubated with T3.70 antibody, followed by FITC-conjugated sheep anti-mouse immunoglobulin antibody, then normal mouse serum and phycoerythrin-labelled anti-CD4 (left column of panels) or biotin-labelled anti-CD8 followed by streptavidin PE (right column of panels). Markers were set arbitrarily and percentages of single positive CD4⁺8⁻ (left column) and CD4⁺8⁺ (right column) expressing $\alpha\beta$ (T3.70⁺) TCR were calculated using FRSP.

Methods. See Fig. 1. For details of staining procedure see Figs 1 and 2 in ref. 6. After incubation with FITC-labelled anti-mouse immunoglobulin antibody and before staining with anti-CD4 or anti-CD8 mAbs, free mouse immunoglobulin binding sites were saturated by incubation with whole mouse serum (2% v/v) for 20 min.

thymocytes (0.4 to 1.6), resulting from an elevated proportion of CD4⁺8⁺ thymocytes, was observed in D^b-expressing recipients of $\alpha\beta$ transgenic stem cells, namely in B6 (B6×DBA/2) F₁ and in HTG mice (Fig. 1 and Table 2). These results show that the preferential differentiation towards CD4⁺8⁺ thymocytes occurs only when transgenic stem cells expressing both α and β transgenic chains develop in the thymic environment expressing restricting D^b MHC molecules.

Staining of thymocytes with the T3.70 antibody demonstrated that cells expressing the highest level of transgenic TCR are virtually absent from D^b-negative recipients (Fig. 1, third row). Double-staining with the T3.70 antibody and anti-CD4 or anti-CD8 antibodies showed that the population expressing high levels of the T3.70 idiotype represents single positive CD4⁺8⁺ thymocytes which have been positively selected in D^b-positive strains, but not in D^b-negative strains (Fig. 2). This observation, the fact that the level of transgenic TCR on CD4⁺8⁺ thymocytes is identical in D^b-positive and D^b negative mice, and the fact that CD4⁺8⁻ thymocytes in all strains do not express high levels of the T3.70 idiotype (Fig. 2), all confirm our earlier conclusions that the CD4/CD8 phenotype of T cells is determined by thymic MHC molecules and the specificity of the TCR on immature thymocytes⁵.

The observation that positive selection of CD4⁺8⁺ thymocytes expressing transgenic TCR occurs in HTG mice and in H-2^b×H-2^d F₁ mice excludes the possibility that the inability to select CD4⁺8⁺ T cells expressing the transgenic D^b-restricted receptor

in non-D^b mice results from suppressive effects of non-D^b loci. Together, the above results indicate that positive selection of H-2D^b-restricted, H-Y antigen-specific T cells in the thymus is strictly dependent on the specific interaction of $\alpha\beta$ TCR with D^b MHC molecules in the absence of nominal antigen. This interaction cannot occur with five other class I MHC molecules expressed in the recipient mice tested in our experiments, that is with K^k, D^k, K^d, D^d and K^b.

Previous experiments were aimed at establishing the selection of the T-cell repertoire by thymic MHC antigens¹²⁻¹⁵. Because the specificity of T cells after *in vivo* immunization was investigated, and not the expression of T-cell receptors on unprimed T cells, these early studies were seen as inconclusive by some authors^{16,17}. We conclude from our experiments that the interaction of the $\alpha\beta$ TCR with thymic MHC antigens determines the CD4/CD8 phenotype of mature T cells, and that this selection exhibits a high degree of specificity. One could argue that the specific low-affinity interaction of TCR with thymic MHC antigens is sufficient to select immature T cells but insufficient to activate mature T cells. The specific low affinity interaction in the thymus may be complemented by accessory molecules present on either immature thymocytes or selecting thymic cells or both. Such a complementation could then lead to the positive selection of T cells with low affinity for self-MHC. Our experiments suggest that all peripheral $\alpha\beta$ T cells have gone through positive selection by thymic MHC antigens, but nevertheless cross-react extensively with allogeneic MHC antigens⁸⁻¹¹. The

selection of allospecific T cells by self-MHC antigens may explain the MHC gene control of responses to allogeneic MHC antigens⁸ and the natural one-way responses of T cells from strains differing in certain MHC antigens⁹. The cross-reactions of self-MHC-restricted T cells with cells expressing allogeneic MHC antigens may be explained by assuming that certain allogeneic MHC antigens bind with higher affinity than self-MHC antigens to the selected receptors. In other cases certain peptides selected by allogeneic MHC antigens may bind with high affinity to selected receptors, even without a significant contribution of the allogeneic MHC antigens to the binding of ligand to the receptor. This may account for the observation of so-called 'allo-MHC-restricted T cells'. Here the selection of a certain peptide rather than the selection of the T-cell receptor

by an allogeneic MHC antigen could result in a specific interaction of a self-MHC-restricted TCR with a target cell expressing the relevant allogeneic MHC antigen plus the selected peptide²⁰⁻²¹. We have previously shown that allo-MHC-restricted T cells tend to cross-react significantly on target cells expressing self-MHC antigens and foreign minor histocompatibility antigens²². These experiments indicate that so-called 'allo-MHC-restricted T cells' were also subject to positive selection by self-thymic MHC antigens.

We thank Katrin Hafen and Verena Stauffer for technical assistance, Dr Werner Haas for discussion and F. Mori for typing the manuscript. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co. Ltd, Basel, Switzerland.

Received 25 July; accepted 12 September 1988.

1. Dembic, Z. *et al.* *Nature* **320**, 232-238 (1986).
2. Saito, T. *et al.* *Nature* **325**, 125-130 (1987).
3. Gabert, J. *et al.* *Cell* **50**, 545-554 (1987).
4. Gay, D. *et al.* *Nature* **328**, 626-629 (1987).
5. Teh, H. S. *et al.* *Nature* **335**, 229-233 (1988).
6. Kisielow, P. *et al.* *Nature* **333**, 742-746 (1988).
7. Staerz, U., Rammensee, H., Benedetto, J. & Bevan, M. J. *Immun.* **134**, 3994-4000 (1985).
8. Bevan, M. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2094-2098 (1977).
9. Lemmonier, F., Burakoff, S. J., Germain, R. N. & Benaceraff, B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1229-1233 (1977).
10. Sredni, B. & Schwartz, R. H. *Immun. Rev.* **54**, 187-223 (1981).
11. von Boehmer, H. *et al.* *Eur. J. Immun.* **9**, 592-597 (1979).
12. Bevan, M. J. *Nature* **269**, 417-419 (1977).
13. Zinkernagel, R. M. *et al.* *J. exp. Med.* **147**, 882-896 (1978).
14. von Boehmer, H., Haas, W. & Jerne, N. K. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2434-2442 (1978).
15. Kappler, J. & Marrack, P. J. *J. exp. Med.* **148**, 1510-1522 (1978).
16. Klein, J. in *Immunology: the Science of Self-Non-Self Discrimination* 267 (Wiley, New York, 1982).
17. Matzinger, P. *Nature* **292**, 497-499 (1981).
18. Egorov, I. K., Mnatsakanyan & Pospelov, L. E. *Immunogenetics* **5**, 65-74 (1977).
19. Meo, T., Vives, J., Miggiano, V. & Shreffler, D. *Trans. Proc.* **5**, 377-381 (1973).
20. Hünig, T. & Bevan, M. J. *J. exp. Med.* **155**, 111-125 (1982).
21. Marrack, P. & Kappler, J. *Nature* **332**, 840-843 (1988).
22. Teh, H. S., Bunnick, J. & von Boehmer, H. *Eur. J. Immun.* **12**, 887-892 (1982).

Spatial distribution of cellular protein binding to retinoic acid in the chick limb bud

M. Maden*, D. E. Ong†, D. Summerbell* & F. Chytil†

* Limb Development Group, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

† Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

Retinoic acid may be the natural morphogen used to generate digit pattern in the chick limb bud¹. It has been proposed that retinoic acid acts by binding to a cellular retinoic acid-binding protein (CRABP) and then entering the nucleus to alter the pattern of gene activity^{2,3}. High-affinity receptors that bind both retinoic acid and DNA and are analogous to the steroid receptors have been identified⁴⁻⁷. But the concentration of endogenous retinoic acid in the limb⁸ and the binding coefficient of the nuclear receptors⁵ indicate that they are saturated throughout the limb. Here we investigate the CRABP distribution in the developing chick limb bud. We find CRABP in the area of intense morphogenetic activity at the tip, with a differential distribution across the anteroposterior axis, the high point being at the anterior margin. Retinoic acid also forms a concentration gradient across the limb bud, but is highest on the posterior side⁸. We propose that CRABP could be reducing the effective concentration of retinoic acid reaching the nucleus to a level appropriate for the differential regulation of gene transcription, providing a spatially modulated morphogenetic gradient of information for digit formation.

When retinoic acid (RA) is applied to the anterior side of the chick limb bud, mirror-symmetrical pattern duplications are produced^{9,10}. The resulting double posterior limbs have up to six digits instead of the normal three. The behaviour of this compound mimics precisely the effects of grafting the zone of polarizing activity, a group of cells at the posterior margin which organizes pattern across the anteroposterior axis of the limb¹¹. This coincidence led to the suggestion that retinoic acid may be the natural morphogen that the zone of polarizing activity uses to generate pattern¹.

We have previously reported that the chick limb bud contains high levels of CRABP at the stages susceptible to pattern

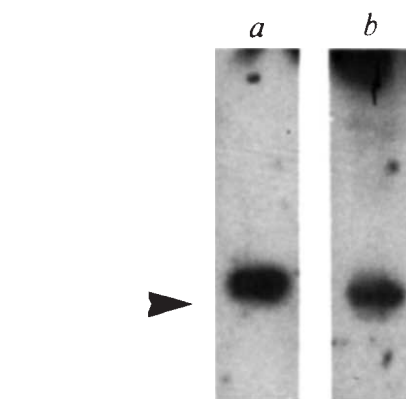
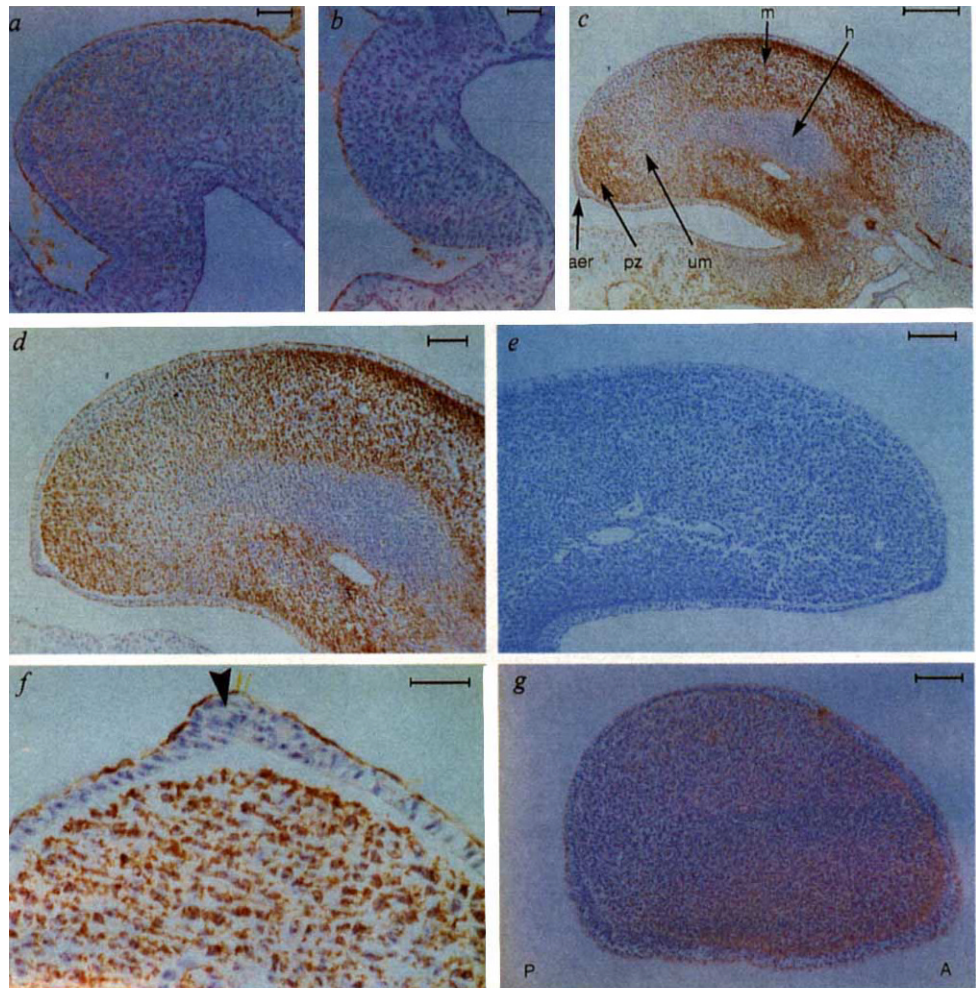


Fig. 1 Detection of immunoreactive proteins in a soluble extract of chick limb buds. *a*, Pure rat CRABP. *b*, Stage 24 chick limb buds. Arrowhead marks the migration of a marker of relative molecular mass 14,400.

Methods. 50-60 stage 24 chick limb buds were homogenized in 100 µl solubilization buffer for SDS-polyacrylamide gel electrophoresis according to Laemmli¹⁸ and centrifuged to remove debris. A 10 µl aliquot of the extract and pure rat CRABP (50 ng) were electrophoresed on an 11% SDS-polyacrylamide gel and the proteins transferred to nitrocellulose paper. Immunoreactivity was detected using an affinity-purified IgG fraction from a rabbit antiserum to rat CRABP and radioiodinated protein A as previously described¹³.

respecification by RA (ref. 12). As in other tissues containing this protein², chick limb-bud CRABP shows a high degree of ligand specificity. The potency of various synthetic and natural retinoids in inducing pattern respecification in the chick limb bud correlates well with their ability to bind CRABP (D.S. and M.M., unpublished observations), suggesting that CRABP may have a role in the mechanism of action of RA in the limb bud. To test this hypothesis further, we used an affinity-purified IgG fraction from a rabbit antiserum to rat CRABP to examine the distribution of CRABP in sections of chick embryos at different stages. The specificity of the staining pattern was established in three ways. First, we demonstrated that the chick limb bud contains a single protein of the same size as rat CRABP on a Western blot (Fig. 1). Second, there was no immunoreactivity on sections using non-immune IgG fractions, and third, we absorbed out the reactivity of the affinity-purified fraction with

Fig. 2 Sections of chick limb buds treated with an affinity-purified rabbit anti-rat CRABP antibody to reveal areas of immunoreactivity. *a*, Central longitudinal section containing the dorsoventral (top to bottom of picture) and proximodistal (right to left) axes of a stage 18 limb bud, showing greater intensity of staining in the distal cells (left of the picture) than in the proximal cells. Bar represents 100 μ m. *b*, Longitudinal section through the posterior margin of the same limb bud as shown in *a*; there is no staining of the cells in this region of the limb bud. Bar represents 100 μ m. *c*, Longitudinal section containing the dorsoventral (top to bottom) and proximodistal (left to right) axes of a stage 24 limb bud. At the tip is the apical ectodermal ridge (aer). Behind this is a dark-staining region that corresponds to the progress zone (pz). Behind that is a weakly staining region consisting of undifferentiated mesoderm (um). Further proximal, the humerus (h) has differentiated in the centre of the limb in the absence of CRABP staining, whereas the muscle, connective tissue and dermis on the periphery (m) stain intensely. Bar, 150 μ m. *d*, Higher-power view of the limb bud in *c*, showing the regionalization of CRABP staining in more detail. Bar, 100 μ m. *e*, Longitudinal section of a limb bud treated as in *c* and *d* except that the affinity-purified fraction was firstly preabsorbed with pure rat CRABP. Absence of staining here confirms the specificity of staining in the other figures. Bar, 100 μ m. *f*, Detail of the lack of staining in the apical ectodermal ridge (arrowhead) and adjacent epidermis, except for the outer periderm layer. Distal mesenchymal cells underneath the ridge stain heavily. Bar, 25 μ m. *g*, Transverse section containing the anteroposterior (left to right) and dorsoventral (top to bottom) axes through the progress zone of a stage-21 limb bud. The anterior margin on the right (A) is far more densely stained than the posterior margin on the left (P). Bar, 100 μ m.



Method. Tissues were fixed in Perfix (Fisher Scientific) for 3 hours, dehydrated, cleared in xylene and embedded in wax. 7 μ m sections were cut. CRABP was immunolocalized with affinity-purified anti-rat CRABP rabbit IgG as previously described, with the exception that the dilution buffer was phosphate buffered saline containing 1% normal goat serum and 0.1% crystalline bovine serum albumin¹³. The IgG fraction was diluted to a calculated absorbance at 280 nm of 0.01. The colour was developed using avidin-biotinylated peroxidase complex (Vector Laboratories, California).

pure rat CRABP (Fig. 2*e*). The isolated IgG fraction has previously been shown not to react with cellular retinol-binding protein¹³.

By stage 18 the limb bud is beginning to emerge from the flank. At this stage three areas of the embryo showed reactivity towards the IgG preparation: peripheral neuroblasts in the spinal cord, cells in the distal regions of the limb bud and the single layer of peridermal cells on the surface of the epidermis. The immunoreactivity in the developing nervous system will be described in detail elsewhere. Concerning the limb bud, it was clear that the cells at the distal tip were the most heavily labelled (Fig. 2*a*). Staining was restricted to these distal limb cells because sections through the posterior margin of the limb bud where it joins the flank revealed little immunoreactivity (compare Fig. 2*a* and *b*). Apart from the single-cell peridermal layer, the remaining layers of the epidermis did not show any reactivity at this stage.

At stage 24, when limb development has progressed to form an elongated bud in which differentiation has begun proximally and distal regions are dividing rapidly, sections show intense staining in a zone at the distal tip (Fig. 2*c* and *d*). This area precisely corresponds to the so-called progress zone: the cells in this location have a high rate of mitosis and are responsible for generating the complex pattern of elements that emerge sequentially during development¹⁴. Only cells within the progress zone are able to respond to the signal from the zone of

polarizing activity, cells in more proximal regions being unaffected¹⁵. Behind the progress zone is an area of undifferentiated mesenchyme that stains less intensely for CRABP (Fig. 2*c* and *d*). In this region cells have begun to withdraw from the cell cycle in preparation for differentiation¹⁴. Behind this zone the beginnings of differentiation can be seen. In the central region the humerus has begun to differentiate in the absence of CRABP reactivity, whereas the muscle and connective tissue in the periphery stain intensely for CRABP. No staining was apparent when the affinity-purified fraction was pretreated with pure rat CRABP (Fig. 2*e*), confirming the specificity of the staining patterns. We have quantified the relative levels of CRABP staining using false colour-image analysis. At stage 18, the ratio of peroxidase label in the cells at the tip of the limb bud to those at the base was 2.3:1. At stage 24 the highest levels were in the proximal periphery (dermis and muscle), followed by distal tip (progress zone), proximal tip (mesenchyme) and proximal core (cartilage) in the ratios 6.1:3.1:1.2:1. The ratio between our lowest signal level and control sections stained at the same time with antibody that had been preabsorbed with pure rat CRABP was 49:1.

The other region of the limb of developmental significance is the apical ectodermal ridge. This structure, like the rest of the epidermis of the embryo, was negative for CRABP staining, apart from the thin peridermal layer on the surface (Fig. 2*f*).

To examine in more detail the staining in the progress zone,

and specifically across the anteroposterior (AP) axis of the limb bud, we cut serial transverse sections of stage 22 limb buds. Sections through the dividing cells of the progress zone revealed that CRABP was differentially distributed across the AP axis of the limb bud (Fig. 2g). The area of highest intensity was at the anterior margin, but staining in the area at the posterior margin was extremely light. The ratio of peroxidase label from anterior to posterior using false colour-image analysis was 3.5:1. It should be stressed that lack of staining cannot be taken to mean the absence of CRABP, only that levels present are below those of the staining areas.

One superficial explanation of the pattern of staining in the limb is that it is caused by local increases in cell density, the highest regions corresponding to the more intensely stained areas. This cannot be the case for several reasons. The progress zone has a similar cell density to the undifferentiated mesenchyme behind it¹⁶, yet has a 2.6-fold greater intensity of stain. The differentiating cartilage has the highest cell density of any area of the limb¹⁶, but is devoid of CRABP staining. Similarly cell counts across the AP axis established that the densities on each side of the limb were similar (posterior, 12.9 cells per mm²; anterior, 14 cells per mm²), yet there is a 3.5-fold difference in staining intensity.

Thus the levels of CRABP in a limb-bud cell during its development seem to be strictly regulated. At birth, CRABP levels are high in the rapidly dividing population of cells in the progress zone. As a cell leaves the progress zone, division slows in preparation for differentiation, and CRABP levels decline. As the developmental decision follows to become cartilage or connective tissue/dermis, CRABP levels are either extinguished (cartilage) or rise again (connective tissue/dermis). These fluctuations suggest that RA and CRABP are important in the development of pattern. Indeed, on the basis of its distribution across the AP axis of the chick limb bud, it has been proposed that RA is the morphogen that organizes pattern in this axis. RA was enriched 2.5-fold in the posterior region compared with the anterior region⁸. It has been calculated that this would result in a change of only 15% in maximum binding across the limb (a rather small difference to specify the complexities of the limb structure), and that an amplification mechanism such as an interaction with CRABP may be present¹⁷. Here we show that the distribution of CRABP is the reciprocal of that of RA, with the anterior region enriched over the posterior region. This would result in an even flatter distribution of the RA/CRABP complex across the limb bud. On the other hand, it would make the distribution of free RA steeper and might help to reconcile the discrepancy between the endogenous concentration of RA (10^{-8} M) and the dissociation constant of the receptor (10^{-10} M– 10^{-9} M; P. Chambon, personal communication). Thus if RA is the morphogen, then CRABP may act to steepen the gradient of morphogen, and the recently identified receptors^{4–7} may be the means by which the pattern of gene transcription is altered.

Increased protein degradation results from elevated free calcium levels found in muscle from *mdx* mice

Paul R. Turner, Tim Westwood, Chester M. Regen & Richard A. Steinhardt*

Department of Zoology, University of California, Berkeley, California 94720, USA

The defective gene responsible for Duchenne muscular dystrophy in humans and the dystrophic condition in *mdx* mice results in a lack of dystrophin at first thought to be localized to the triads, but more recently found on the cytoplasmic side of the sarcolemma of skeletal muscle fibres^{1–5}. Because the total calcium content of dystrophic fibres is significantly raised^{6–9}, we have compared the intracellular free calcium concentration ($[Ca^{2+}]_i$) in skeletal muscle in normal and *mdx* mice. We find that $[Ca^{2+}]_i$ is markedly elevated in *mdx* muscle fibres compared with normal fibres, both at rest and during stimulation. By measuring protein degradation rates and manipulating $[Ca^{2+}]_i$, we have been able to demonstrate directly that the elevation of $[Ca^{2+}]_i$ in *mdx* fibres results in an enhanced net degradation of muscle proteins.

Resting levels of $[Ca^{2+}]_i$ were measured in individual single fibres of the flexor digitorum brevis muscle shortly after microinjection with the fluorescent tetracarboxylate chelator fura-2. The value obtained from 24 fibres from normal adults and 22 *mdx* fibres, with an external calcium ion concentration ($[Ca^{2+}]_o$) of 1.84 mM, was $40 \text{ nM} \pm 2.8$ (mean $\pm$ standard error of the mean) in normal mice fibres and $92 \pm 9.8 \text{ nM}$ in *mdx* fibres (Fig. 1). An elevated $[Ca^{2+}]_i$ has recently been reported for cultured human dystrophic muscle cells, using the fluorescent tetracarboxylate chelator quin-2¹⁰. These results indicate that dystrophic fibres have a reduced ability to regulate $[Ca^{2+}]_i$. To test this hypothesis, we examined the response of skeletal muscle fibres from normal and *mdx* mice to varied extracellular calcium levels. When $[Ca^{2+}]_o$ was reduced to 0.184 mM, $[Ca^{2+}]_i$ was lowered by 40% in *mdx* fibres to a value close to the normal resting level (Fig. 1). Under identical conditions $[Ca^{2+}]_i$ fell by only 15% in normal fibres (Fig. 1). $[Ca^{2+}]_i$ therefore varies more dramatically with $[Ca^{2+}]_o$ in *mdx* muscle, especially in the range of the normal $[Ca^{2+}]_o$ of 1.84 mM (Fig. 1). These results are consistent with the hypothesis that the site of the primary defect in dystrophic muscle is the sarcolemma.

From reports that dystrophin was located in the triad structure in muscle fibres^{2,5} and from observations on calcium sequestration rates in *mdx* muscle¹¹, it seemed possible that the site of the defect could be the sarcoplasmic reticulum (SR). To examine the kinetics of calcium release and uptake by the SR, individual fura-2 ester-loaded fibres were stimulated electrically using platinum electrodes. We used the ester-loading method in these experiments to obtain more successful loadings for each preparation, because it was not always possible to arrange stimulation for a single injected fibre, and we were not as concerned with absolute values for $[Ca^{2+}]_i$. The fluorescence at 500–530 nm resulting from alternating periods of excitation at 350 nm and 385 nm was recorded in 100 μs bins, for periods of 75–150 ms after the stimulus. Fibres were stimulated at various frequencies (0.5–9.8 Hz), and 10–50 ratios were averaged for each point in a record (Fig. 2).

At 37 °C the calcium transient peak levels following stimulation were similar in normal and *mdx* fibres (Figs. 2 and 3a). For a given fibre, stimulation at higher frequencies resulted in a drop in peak amplitude (data not shown) but peak values were maximal at stimulus rates of 3.3 Hz or less. The mean time

Received 28 July; accepted 16 September 1988.

- Eichele, G., Tickle, C. & Alberts, B. M. *J. Cell Biol.* **101**, 1913–1920 (1985).
- Chytil, F. & Ong, D. A. in *The Retinoids* (eds Sporn, M. B., Roberts, A. B. & Goodman, D. S.) 89–123 (Academic, Orlando, Florida, 1984).
- Takase, S., Ong, D. E. & Chytil, F. *Archs. Biochem. Biophys.* **247**, 328–334 (1986).
- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444–450 (1987).
- Giguere, V., Ong, D. E., Segui, P. & Evans, R. M. *Nature* **330**, 624–629 (1987).
- Brand, N. *et al.* *Nature* **332**, 850–853 (1988).
- Benbrook, D., Lernhardt, E. & Pfahl, M. *Nature* **333**, 669–672 (1988).
- Thaller, C. & Eichele, G. *Nature* **327**, 635–628 (1987).
- Tickle, C., Alberts, B., Wolpert, L. & Lee, J. *Nature* **296**, 564–566 (1982).
- Summerbell, D. *J. Embryol. exp. Morph.* **78**, 269–289 (1983).
- Tickle, C., Summerbell, D. & Wolpert, L. *Nature* **254**, 199–202 (1975).
- Maden, M. & Summerbell, D. *J. Embryol. exp. Morph.* **97**, 239–250 (1986).
- Porter, S. B., Ong, D. E. & Chytil, F. *J. Androl.* **6**, 197–212 (1985).
- Wolpert, L., Lewis, J. & Summerbell, D. in *Cell Patterning* (Ciba Found. Symp. **29**) 95–130 (Associated Scientific, Amsterdam, 1975).
- Summerbell, D. *J. Embryol. exp. Morph.* **32**, 227–237 (1974).
- Summerbell, D. & Wolpert, L. *Nature new Biol.* **239**, 24–26 (1972).
- Robertson, M. *Nature* **330**, 420–421 (1987).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).

* To whom correspondence should be addressed.

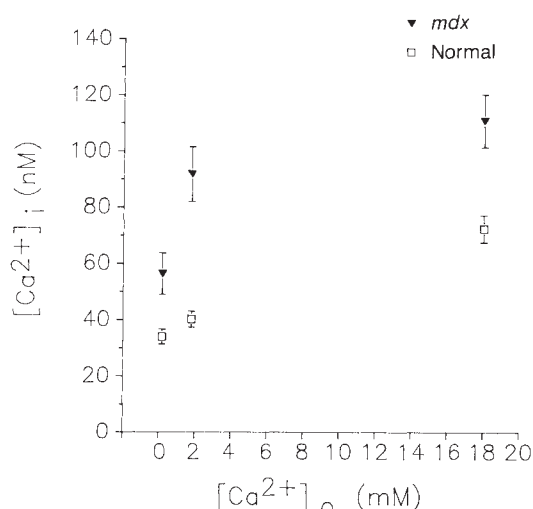


Fig. 1 The resting free calcium levels in skeletal muscle fibres from normal mice (squares) and *mdx* dystrophic mice (triangles), in Ringer solutions with low (0.184 mM) calcium, normal (1.84 mM) calcium and high (18.4 mM) calcium. Male mice used were 1–4 months old. Bars indicate standard errors. All experiments were at 37 °C. The ratios used to determine resting levels were obtained shortly after dye injection¹⁶. In those experiments where $[Ca^{2+}]_o$ was varied, the ratio was allowed to reach a new resting value, for which an equilibration period of several minutes was often required. Mean resting $[Ca^{2+}]_i$ values are: at low calcium: 56.4 ± 7.4 nM in *mdx* fibres ($n = 10$) and 34.1 ± 2.6 nM in normal fibres ($n = 10$); at normal calcium: 92 ± 9.8 nM in *mdx* fibres ($n = 22$) and 40 ± 2.8 nM in normal fibres ($n = 24$); and at high calcium: 111 ± 9.4 nM in *mdx* fibres ($n = 7$) and 72 ± 4.8 nM in normal fibres ($n = 8$).

Methods. The flexor digitorum brevis muscle was dissected intact from the mouse hind foot and pinned to a circle of silicone adhesive on a glass bottom of a Petri dish. Fibres were stretched to sarcomere lengths of about 2–4 μ m (resting length). The mouse Ringer solution used contained the following (in mM): NaCl, 138; KCl, 2.7; $CaCl_2$, 1.84; $MgCl_2$, 1.06; glucose, 5.6; HEPES, 12.4 pH 7.20. The resting potentials of fibres after these procedures were measured using electrodes filled with 3M KCl (electrode resistance was 20–60 M Ω , tip potentials ≤ 5 mV) and were -76 ± 18 mV ($n = 7$) for normal fibres, and -68 ± 10 mV ($n = 6$) for *mdx* fibres. These were slightly more negative than resting potentials previously reported for single dissociated fibres of this muscle type isolated from adult rats¹⁷. Fibres were viewed through a 10 $\times$ objective. The background fluorescence intensities at 500–530 nm resulting from excitation at 350 nm and 385 nm were recorded before injection. Fibres were microinjected with a solution of 27.5 mM fura-2 (Molecular Probes), 66 mM potassium gluconate, 5 mM $CaCl_2$, 5 mM HEPES, pH 7.3¹⁶. Injection volumes were $\leq 1\%$ of fibre volume (≤ 14 pl)¹⁸. After injection the ratio was typically slightly raised, but fell rapidly as the fibre recovered; values were recorded after this recovery period. Free calcium was calculated from the fluorescence ratio (R) obtained by dividing the fluorescence (500–530 nm) values from excitation at 350 nm by those at 385 nm¹⁹. It was necessary to correct the ratios obtained in calibration solutions to determine $[Ca^{2+}]_i$ values²⁰ because of increased dye fluorescence inside the fibres at 385 nm excitation, presumably because of differences in intracellular viscosity, ionic strength, and/or binding of dye²¹. The correction factor was determined by incubation of fura-2-injected fibres in the presence of 10 μ M ionomycin with various $[Ca^{2+}]_o$ ²², the peak ratio before loss of dye being used, and by injection of fibres with buffered calcium/EGTA solutions such that $[Ca^{2+}]_i$ within the fibre was calculated to be 80 or 140 nM¹⁸. The correction factor obtained from the two types of calibration experiment was identical in both normal and dystrophic fibres (0.75 ± 0.03).

to reach the peak value was also not significantly different at 37 °C (Fig. 3b). However, the half-time for recovery to resting levels ($T_{1/2}$) from transient peaks seemed noticeably longer in *mdx* fibres (Fig. 3c). When fibres were being repeatedly stimulated, the stimulated 'resting' or basal $[Ca^{2+}]_i$ was markedly elevated (Fig. 2). This basal elevation was dependent on the

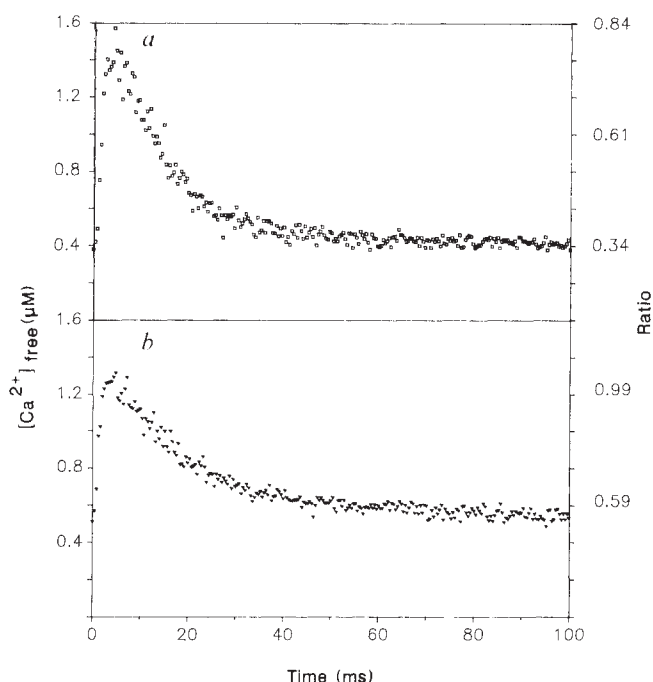


Fig. 2 Calcium transients in normal (a) and dystrophic (b) single fibres, following electrical stimulation at 37 °C. Each point represents the average of 50 (normal) or 30 (*mdx*) 100 μ s bins at 350 nm divided by the average of 50 (normal) or 30 (*mdx*) 100 μ s bins at 385 nm.

Methods. Fibres were dissected and pinned as described above, and then loaded with ester by mixing equal volumes of 10 mM fura-2 acetmethoxy (AM) ester in dimethyl sulphoxide and 25% w/w Pluronic F-127 (BASF Wyandotte, Michigan) and by diluting this to a final concentration of 10 μ M fura-2 AM ester in the dish. Muscles were incubated in fura-2 AM at 25 °C for 60 min, after which the muscles were rinsed several times in fura-2 free ringer, and ratio measurements were made at 37 °C or 25 °C shortly thereafter²⁰. A significant elevation of $[Ca^{2+}]_i$ in *mdx* fibres was also found using the AM-ester dye-loading method: at 37 °C normal fibres were found to have a mean resting $[Ca^{2+}]_i$ of 201 ± 6 nM, $n = 18$; and *mdx* fibres a mean $[Ca^{2+}]_i$ of 282 ± 13 nM, $n = 28$ ($P < 0.001$). The higher values obtained from ester loading, although comparable to those seen in smooth muscle²² and cardiac muscle²³ with fura-2 AM ester, may reflect the presence of a calcium-insensitive component of the dye²⁴, or significant compartmentalization of the dye²⁰. A correction factor of 0.33 was determined for fura-2 loaded by the AM ester method²². Fibres were stimulated at a rate of 3.3 Hz using platinum electrodes placed close to the two sides of the fibre. The data acquisition computer was used to trigger a stimulator that sent brief depolarizing pulses to the electrodes. Data collection began 100 μ s after each stimulus. A slit aperture was placed over each fibre to exclude any fluorescence from adjacent fibres. Movement artefacts were sometimes observed in the intensity recordings, but were cancelled out by the ratio method¹⁹. It was not possible to obtain background readings for the fibres that were stimulated, and backgrounds were not subtracted from the recordings. Backgrounds of fibres before loading with fura-2 AM ester were obtained however, and the lack of subtraction should result in a mean error of $3.12 \pm 0.45\%$ ($n = 15$) in the ratio, or $< 5\%$ in the $[Ca^{2+}]_i$ values. The fibres were examined visually to ascertain that they were twitching in response to the stimulating frequency (which could be followed visually up to ~ 10 Hz). Sarcomere spacings were: a, 2.9 μ m and b, 3.1 μ m. The apparent resting $[Ca^{2+}]_i$ values were elevated during stimulation: at 37 °C, at rates of ≤ 3.3 Hz, the mean $[Ca^{2+}]_i$ rose to 353 ± 20 nM ($n = 6$) in normal fibres, and to 424 ± 15 nM ($n = 3$) in *mdx* fibres. This difference in mean $[Ca^{2+}]_i$ was significant ($P < 0.01$). At 25 °C, the mean $[Ca^{2+}]_i$ rose from 123 ± 4 nM to 228 ± 14 nM ($n = 17$) in normal fibres when stimulating, and from 171 ± 7 nM to 290 ± 22 nM ($n = 23$) in *mdx* fibres ($P < 0.01$).

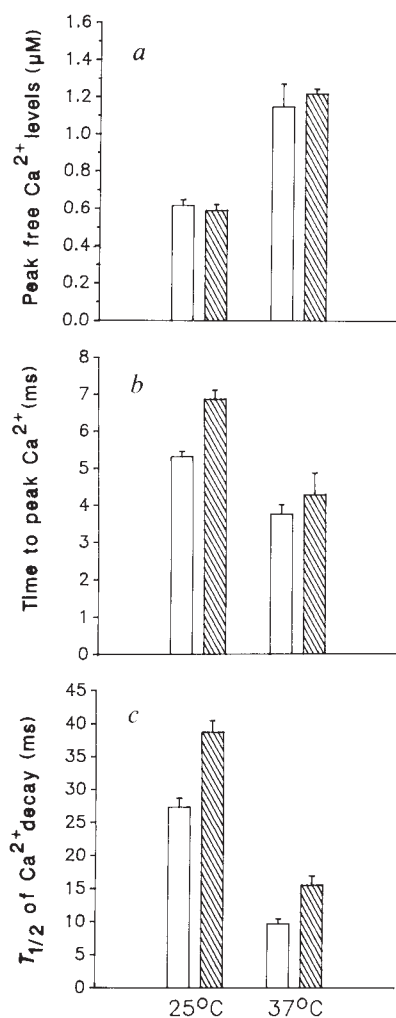


Fig. 3 Summary of calcium transient data from normal (open) and dystrophic muscle fibres (cross-hatched). The mean results at 25 °C from 17 normal fibres (310 ratios) and 23 *mdx* fibres (530 ratios), and at 37 °C from six normal fibres (130 ratios) and three *mdx* fibres (60 ratios) are shown. Time to peak values and $T_{1/2}$ values were similar to those seen in rat skeletal muscle fibres²⁵. **a**, Peak $[Ca^{2+}]_i$ levels (μM) obtained at 25 °C and 37 °C, following stimulation. At 25 °C: normal mean peak is $0.617 \pm 0.030 \mu M$, *mdx* mean peak is $0.589 \pm 0.32 \mu M$; difference was not significant ($P > 0.10$). At 37 °C: normal mean peak is $1.148 \pm 0.123 \mu M$, *mdx* mean peak is $1.217 \pm 0.026 \mu M$; $P > 0.10$. Stimulus frequencies were ≤ 3.3 Hz (peak values were maximal). **b**, The mean time for the transient to reach a peak value at 25 °C and 37 °C, in ms. At 25 °C: normal mean is 5.31 ± 0.14 ms, *mdx* mean is 6.88 ± 0.25 ms; $P < 0.001$ (highly significant). At 37 °C: normal mean is 3.75 ± 0.26 ms, *mdx* mean is 4.27 ± 0.58 ms; a difference that was not significant ($P > 0.10$). **c**, The mean half-decay times ($T_{1/2}$) for the calcium transients return to elevated resting levels, at 25 °C and 37 °C, in msec. At 25 °C: normal mean $T_{1/2}$ is 27.28 ± 1.38 ms, *mdx* mean $T_{1/2}$ is 38.71 ± 1.73 ms; $P < 0.001$ (highly significant). At 37 °C: normal mean $T_{1/2}$ is 9.70 ± 0.70 ms, and *mdx* mean $T_{1/2}$ is 15.53 ± 1.39 ms; $P < 0.05$ (significant).

frequency of stimulation, being greater at 9.8 Hz than at 3.3 Hz (data not shown), with the elevated basal $[Ca^{2+}]_i$ values always significantly higher in *mdx* fibres than in normal fibres (Fig. 2). Although peak $[Ca^{2+}]_i$ values were identical for *mdx* and normal fibres at 25 °C, albeit lower than at 37 °C (Fig. 3a), the differences in transient kinetics seen at 37 °C were enhanced at this lower temperature. The time taken to reach peak $[Ca^{2+}]_i$ was significantly longer at 25 °C, (Fig. 3b); and the $T_{1/2}$ was again significantly prolonged in *mdx* fibres (Fig. 3c).

Thus, although peak $[Ca^{2+}]_i$ values were similar in normal and *mdx* fibres, because of the slowed calcium transient kinetics,

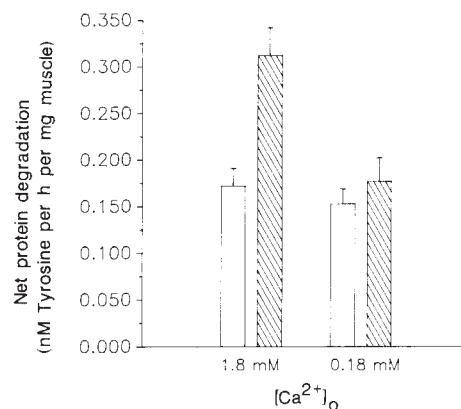


Fig. 4 Rates of net protein degradation in normal (open) and *mdx* (cross-hatched) soleus muscles, in normal (1.84 mM) and low (0.184 mM) extracellular calcium. Each bar represents the mean of four experiments $\pm$ standard error of the mean. Male mice used were 3–6 weeks old. The *mdx* rate in normal (1.84 mM Ca^{2+}) calcium, 0.312 ± 0.030 nM tyrosine $h^{-1} mg^{-1}$ is higher than the rate in normal muscle, 0.172 ± 0.019 nM $h^{-1} mg^{-1}$; $P \leq 0.01$. The *mdx* rate in low (0.184 mM) calcium was 0.177 ± 0.025 nM $h^{-1} mg^{-1}$, the normal rate was 0.153 ± 0.016 nM $h^{-1} mg^{-1}$. The normal muscle degradation rates were similar to those found previously in rat soleus muscles^{12–15}.

Methods. Soleus muscles were dissected intact, blotted dry and weighed. Muscles were pinned at resting lengths and incubated at 37 °C in mouse Ringer (as above) supplemented with 0.85 mM leucine, 0.5 mM isoleucine, 1.0 mM valine and 0.1 U ml^{-1} insulin¹⁴. Two hours later the medium was removed and assayed for tyrosine¹³, and the amount of tyrosine per h per mg muscle wet weight determined. Muscles were homogenized in 5% trichloroacetic acid to determine the intracellular tyrosine pool sizes¹⁴. Pool sizes were identical for *mdx* and normal muscles in both low and normal $[Ca^{2+}]_i$; mean size being 0.077 ± 0.009 nM per mg muscle, and did not vary during these experiments.

$[Ca^{2+}]_i$ levels remain elevated significantly longer per twitch in *mdx* fibres stimulated at physiological frequencies. In addition, the basal $[Ca^{2+}]_i$ is elevated to a greater extent in *mdx* fibres during stimulation (Fig. 2). The slowed transient kinetics in *mdx* fibres could be a result of the elevation in resting $[Ca^{2+}]_i$. It is also conceivable that sarcolemmal calcium fluxes could be a component of the *mdx* calcium transients, so these results are not inconsistent with a defect in the sarcolemma.

It has been proposed that an elevation in $[Ca^{2+}]_i$ could promote net muscle protein degradation, by directly or indirectly stimulating protease activity^{9,12–15}. We tested this hypothesis by comparing rates of protein degradation in normal and *mdx* muscle, while varying $[Ca^{2+}]_o$ to manipulate $[Ca^{2+}]_i$ (see Fig. 1). The release of tyrosine, which is not metabolized by muscle, was used as an indicator of net protein degradation^{13,14}. The soleus muscle was used because normal degradation rates have been previously measured in this muscle^{12–15}, and because the mouse flexor digitorum brevis muscle is small.

When degradation was measured under conditions of normal $[Ca^{2+}]_o$ of 1.84 mM, dystrophic muscles had a significantly higher rate of degradation: 80% higher than in normal muscle (Fig. 4). When $[Ca^{2+}]_o$ was reduced 10-fold to 0.184 mM, the rate of degradation in normal muscle was lowered by only 10% (Fig. 4), a result consistent with the small changes in $[Ca^{2+}]_i$ seen under these conditions (Fig. 1). Thus net protein degradation in normal muscle, maintained under tension in the presence of branched-chain amino acids and insulin, seems to be relatively independent of $[Ca^{2+}]_o$ ¹⁵. But, in dystrophic muscle with a $[Ca^{2+}]_o$ of 0.184 mM, the degradation rate fell dramatically to a level close to that seen in normal muscle with normal $[Ca^{2+}]_o$ (Fig. 4). Under these conditions, $[Ca^{2+}]_i$ in *mdx* fibres was close to normal fibre values (Fig. 1), and the concomitant drop in the *mdx* degradation rate to the normal level (Fig. 4) demonstrates that the higher degradation rate seen in *mdx*

muscle was the result of the higher intracellular free calcium. It has been shown previously that proteolysis in cut rat diaphragm muscles is also particularly sensitive to $[Ca^{2+}]_o$, which is presumably due to an increased permeability of those cut fibres to $[Ca^{2+}]_o$ ¹⁵. We have, in this study, directly compared the intracellular levels of free calcium and protein degradation rates.

The absence of dystrophin, directly or indirectly, results in the inability of *mdx* fibres to maintain a low $[Ca^{2+}]_i$ at rest and also results in a prolonged elevation of $[Ca^{2+}]_i$ during stimulation. This elevation induces a significant increase in the net degradation of muscle proteins.

We thank Richard Strohman for discussions and for the mice used in this study, and Woody Hopf for assistance with the protein degradation experiments. This work was supported in part by the NIH (R.S.) and the Muscular Dystrophy Association (P.T.). IBM donated the computer.

Received 14 July; accepted 16 September 1988.

- Hoffman, E. P., Brown, R. H. Jr. & Kunkel, L. M. *Cell* **51**, 919-928 (1987).
- Hoffman, E. P., Kudson, C. M., Cambell, K. P. & Kunkel, L. M. *Nature* **330**, 754-757 (1987).
- Sugita, H. *et al. Proc. Japan Acad.* **64B**, 37-39 (1988).
- Zubrzycka-Gaarn, E. E. *et al. Nature* **333**, 466-469 (1988).
- Watkins, S. C., Hoffman, E. P., Slayter, H. S. & Kunkel, L. M. *Nature* **333**, 863-866 (1988).
- Bertorini, T. E. *et al. Neurology* **32**, 1088-1092 (1982).
- Bertorini, T. E. *et al. Neurology* **34**, 1436-1440 (1984).
- Jackson, M. J., Jones, D. A. & Edwards, R. H. T. *Clin. chim. Acta* **147**, 215-222 (1985).
- Duncan, C. J. *Experimenta* **34**, 1531-1535 (1978).
- Mongini, T. *et al. Neurology* **38**, 476-480 (1988).
- Mobley, B. A. *Exp. Neurol.* **87**, 137-146 (1985).
- Zeman, R. J., Kameyama, T., Matsumoto, K., Bernstein, P. & Etlinger, J. D. *J. biol. Chem.* **260**, 13619-13624 (1985).
- Fulks, R. M., Li, J. B. & Goldberg, A. L. *J. biol. Chem.* **250**, 290-298 (1975).
- Furono, K. & Goldberg, A. L. *Biochem. J.* **237**, 859-864 (1986).
- Baracos, V., Greenberg, R. E. & Goldberg, A. L. *Am. J. Physiol.* **250**, E702-710 (1986).
- Poenie, M., Alderton, J., Tsien, R. Y. & Steinhardt, R. A. *Nature* **315**, 147-149 (1985).
- Bekoff, A. & Betz, W. J. *Physiol.* **271**, 537-547 (1977).
- Turner, P. R., Jaffe, L. A. & Fein, A. J. *Cell Biol.* **102**, 70-76 (1986).
- Gryniewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440-3450 (1985).
- Poenie, M., Alderton, J., Steinhardt, R. & Tsien, R. *Science* **223**, 886-889 (1986).
- Tsien, R. Y., Rink, T. F. & Poenie, M. *Cell Calcium* **6**, 145-157 (1985).
- Williams, D. A., Fogarty, K. E., Tsien, R. Y. & Fay, F. S. *Nature* **318**, 558-561 (1985).
- Wier, W. G., Cannell, M. B., Berlin, J. R., Marban, E. & Lederer, W. J. *Science* **235**, 325-328 (1987).
- Scanlon, M., Williams, D. A. & Fay, F. S. *J. biol. Chem.* **262**, 6308-6312 (1987).
- Eusebi, F., Miledi, R. & Takahashi, T. *Nature* **284**, 560-561 (1980).

Functional replacement of the HIV-1 rev protein by the HTLV-1 rex protein

Laurence Rimsky, Joachim Hauber, Mitchell Dukovich, Michael H. Malim, Alphonse Langlois*, Bryan R. Cullen & Warner C. Greene

Howard Hughes Medical Institute, Departments of Medicine, Microbiology and Immunology and *Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA

Two evolutionarily distinct families of human retroviruses, the human immunodeficiency viruses (HIV) and the human T-cell leukaemia viruses (HTLV), have been defined (reviewed in ref. 1). Although these virus groups share tropism for human CD4⁺ T cells, they differ markedly in primary sequence, genetic organization and disease association (AIDS versus adult T-cell leukaemia), but show similar general strategies for the regulation of viral gene expression. Each encodes a protein able to *trans*-activate transcription from the homologous viral long terminal repeat (tat in HIV^{2,3}, tax in HTLV⁴⁻⁷), although these proteins act by different mechanisms and do not appear to be interchangeable⁸⁻¹¹. Each virus also produces a second *trans*-acting protein that induces the expression of the unspliced messenger RNAs encoding the viral structural proteins (rev in HIV^{12,13} and rex in HTLV¹⁴). Here we show that the rex protein of HTLV-I can functionally replace the rev protein of HIV-1 in transient expression assays. This genetic complementation by rex is adequate for the rescue of a replication-defective

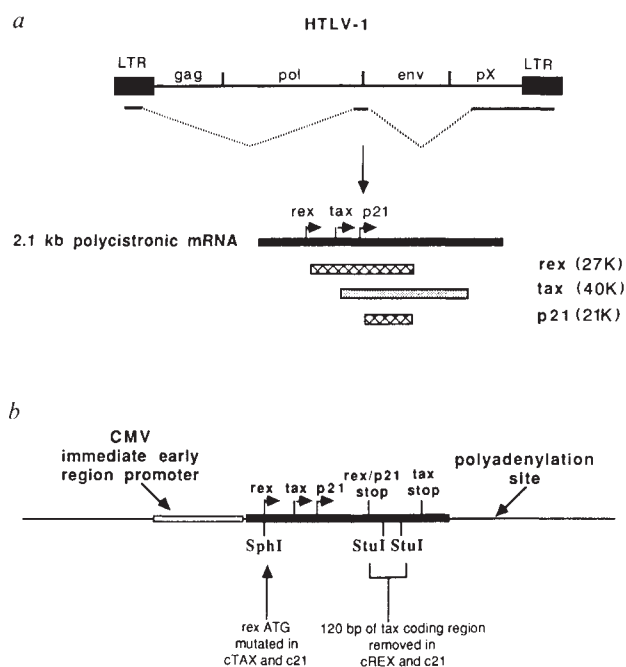


Fig. 1 Overview of rex production by HTLV-I and strategy for the construction of the rex, tax and p21 cDNA expression vectors. *a*, A double-splicing event involving sequences within the *pol-env* and *pX* regions of the HTLV-I provirus leads to the production of a subgenomic, polycistronic mRNA encoding the 27K rex protein, the 40K tax protein, and a 21K protein referred to as p21 (refs 17, 27). The relative position of the start sites of translation for these three proteins is indicated by the arrows. Rex and p21 are encoded within the same reading frame and differ only in an additional 78 N-terminal amino acids present in the rex protein. The tax protein is translated from a different reading frame and so does not share significant amino-acid sequence similarity with rex or p21. *b*, Construction of pcREX, pcTAX and pc21 cDNA expression vectors. A polycistronic pX cDNA (as depicted in *a*) was isolated from human T cells infected with the C91/PL strain of HTLV-I. This cDNA, or modified versions detailed below, was inserted into the expression vector pBC12/CMV (ref. 9). For the preparation of pcREX, a 120-base pair *StuI-StuI* fragment was removed that corresponded to a portion of the tax protein not present in the rex or p21 proteins. This vector encodes rex and p21 but not tax, as deduced by an inability to mediate transactivation of the HTLV-I LTR or to direct the synthesis of a protein immunoprecipitated by a tax-specific heteroantibody (data not shown). The pcTAX expression vector encoding tax and p21, but not rex, was prepared by site-directed deletional mutagenesis of the rex translation initiation codon, which is coincident with the indicated *SphI* site. The pc21 expression vector, which encodes p21 but not tax and rex, was prepared by deletion of both the 120-base pair *StuI-StuI* fragment of *tax* and the initiation codon for *rex*. These plasmids were introduced into monkey kidney COS cells using a DEAE-dextran transfection procedure specifically optimized for this cell line⁹. Data represent results from three to eight different experiments involving at least two independent preparations of each expression vector.

rev mutant of HIV-1. This unexpected shared function between the structurally distinct rex and rev proteins emphasizes the importance of this highly conserved pathway for the regulation of human retrovirus gene expression.

The *rev*-encoded protein of HIV-1 and the *rex*-encoded protein of HTLV-I are each obligately needed for viral replication and are specifically required for the functional expression of unspliced (or singly spliced) mRNAs encoding the *gag*, *pol* and *env* structural gene products¹²⁻¹⁶. Each of these regulatory proteins appears to act after transcription, possibly by directly suppressing RNA splicing or enhancing the nuclear export of these larger viral mRNAs. The HTLV-1 rex protein is normally

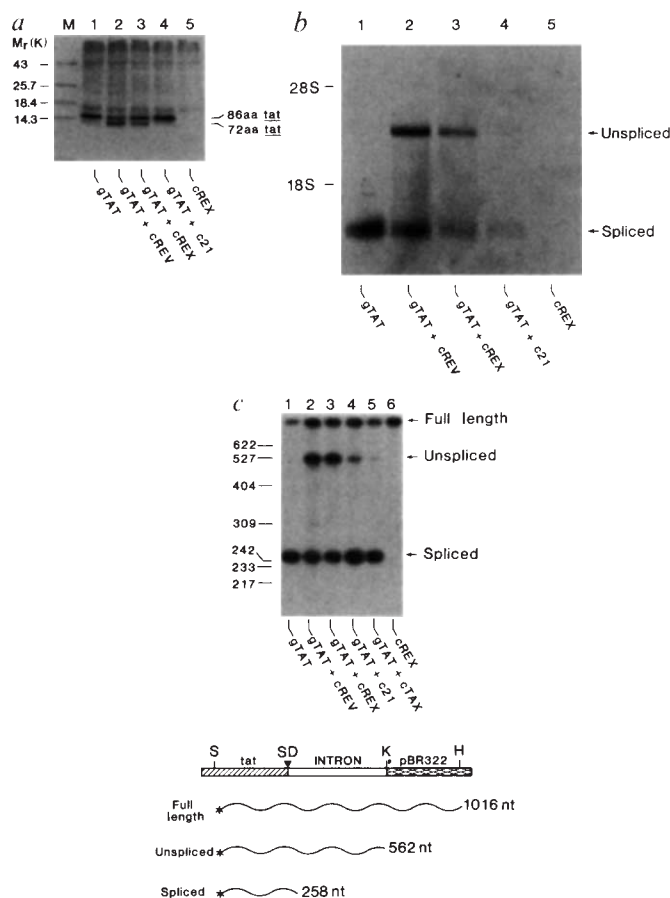


Fig. 2 Comparison of HIV-1 rev and HTLV-I rex effects on the pattern of HIV-1 tat protein and mRNA expression. COS cells were cotransfected with the pgTAT expression vector, which contains a genomic version of the HIV-1 *tat* gene (first and second coding exons and intervening intron) but lacks the capacity to encode rev (see ref. 18 for detailed description) and the pcREV, pcREX, pc21 or pcTAX expression vectors. Some cells were transfected with pcREV in the absence of pgTAT. **a**, Anti-tat immunoprecipitation²⁸ of [³⁵S]cysteine radiolabelled extracts prepared from cells transfected with the indicated plasmids. The full-length (86 amino acids) and truncated (72 amino acids) forms of the tat protein are identified and the sizes of calibration markers (lane M) are indicated. **b**, Northern blot analysis of cytoplasmic RNA²⁹ isolated from transfected cells hybridized with a radio-labelled *tat*-specific DNA probe. Spliced and unspliced forms of the tat mRNA are indicated by arrows and the relative migration of the 28S and 18S ribosomal RNAs is also shown. **c**, S1 nuclease protection analysis of cytoplasmic RNA isolated from COS cells transfected with the indicated plasmids. Schematic indicates the 5' end-labelled probe used and the predicted size of fragments protected by spliced (258 nucleotides) and unspliced (562 nucleotides) forms of tat mRNA.

Methods. COS cells were cotransfected ($1.25 \mu\text{g ml}^{-1}$ of each plasmid) with pgTAT+pBC12/CMV/IL-2 (negative control), pgTAT+pcREV, pgTAT+pcREX, pgTAT+pc21, pgTAT+pcTAX or pcREV+pBC12/CMV/IL-2. 60 Hours later, cells were radiolabelled with [³⁵S]cysteine for 2 h, extracted in nonionic detergent and immunoprecipitated^{18,28}. Immunoprecipitates were analysed on 14% SDS-polyacrylamide gels electrophoresed under reducing conditions (**a**). In other experiments, cytoplasmic mRNA was prepared from transfected cells²⁴ at 60 hours and analysed by Northern blotting using a *tat*-specific probe (*SalI*-*MstII* fragment) (**b**) or by S1 nuclease protection with the indicated probe⁹ (**c**).

translated from a subgenomic mRNA containing sequences derived in part from the pX region of this virus¹⁷ (Fig. 1a). This rex mRNA is polycistronic and also encodes the 40K tax protein (relative molecular mass M_r 40,000) from a different reading frame and a 21K protein (p21) from the same reading frame. The p21 protein lacks 78 N-terminal amino acids present in the rex protein (189 amino acids). To identify any functional similarities in the HTLV-I rex and HIV-1 rev proteins, we prepared a matched series of HTLV-I and HIV-1 complementary DNA expression vectors (Fig. 1b). The HTLV-I vectors and their respective gene products included: pcREV (rex, p21), pcTAX (tax and p21), pc21 (p21 only). The previously described pcREV vector¹⁸ was used for expression of the HIV-1 rev protein.

As well as initiating the synthesis of viral structural proteins, the rev protein induces the production of a truncated form of the HIV-1 tat protein¹⁸. This foreshortened tat protein is produced by translation of an unspliced form of tat mRNA which retains a highly conserved stop codon located immediately downstream from the normal splice donor site within the *tat* intron. Sequences within this *tat* intron coincide with the *env* gene coding region. Thus, rev induction of the smaller form of the *tat*-encoded protein reflects the generation of an unspliced mRNA encoding the virion envelope protein. We initially exploited this rev-induced change in the size of the HIV-1 tat protein to investigate whether the rex protein of HTLV-I could functionally mimic the action of the rev protein. Transfection of COS cells with an expression vector encoding a genomic version of *tat*, pgTAT (ref. 18), led exclusively to the production of the full-length 86-amino-acid form of the HIV-1 tat protein encoded by the spliced *tat* mRNA (Fig. 2a, lane 1). As we have previously described¹⁸, cotransfection of the pcREV expression vector with pgTAT resulted in the appearance of the smaller 72-amino-acid form of the tat protein encoded by the unspliced *tat* mRNA (lane 2). Strikingly, transfection of the pcREX expression vector encoding the rex and p21 proteins of HTLV-I

also gave substantial amounts of the 72-amino-acid version of the tat protein (lane 3). But the pc21 expression vector did not detectably induce the production of this smaller tat protein, indicating that rex was the biologically active protein species encoded by the pcREX vector (lane 4). As expected, neither form of the tat protein was identified in lysates of cells transfected with the pcREV vector alone (lane 5).

To confirm that the rex protein induced the production of the larger unspliced form of tat mRNA that was predicted to encode the smaller tat protein, we hybridized a Northern blot of cytoplasmic RNA from transfected cells with a *tat*-specific DNA probe (Fig. 2b). Transfection of COS cells with the pgTAT vector alone yielded a single cytoplasmic RNA species of the size predicted for fully spliced *tat* mRNA (lane 1). But cotransfection of either the rev or rex expression vectors with pgTAT gave a larger RNA of the size predicted for unspliced *tat* mRNA (lanes 2 and 3). Transfection of the pc21 vector has no obvious effect (lane 4), although the presence of lesser amounts of spliced *tat* mRNA precluded firm conclusions. The *tat* specificity of the probe was confirmed by the lack of a detectable hybridization signal in cells transfected with pcREV alone (lane 5).

We performed S1 nuclease protection assays to show that the larger cytoplasmic *tat* mRNAs detectable in the presence of either the rev or rex proteins were indicative of a diminished use of the splice-donor site located at the end of the first coding exon of *tat* (Fig. 2c). Cytoplasmic RNA from COS cells transfected with pgTAT alone protected a 258-nucleotide probe fragment, showing that these cells contained only fully spliced *tat* mRNA (lane 1). In contrast, cytoplasmic RNA from cells cotransfected with pgTAT and either pcREV or pcREX gave a second protected band of 562 nucleotides, a size consistent with that predicted for the unspliced form of *tat* mRNA (lanes 2 and 3). Interestingly, this sensitive assay also revealed the presence of small amounts of unspliced mRNA in cells transfected with

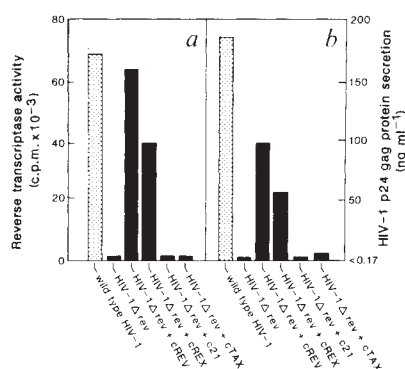


Fig. 3 The rex protein of HTLV-I rescues the replication of a defective HIV-1 provirus mutated in the *rev* gene. COS cells were cotransfected with either a wild-type HIV-1 provirus (pHXB2gpt, ref. 13), a *rev* mutant of this virus (HIV-1 Δ *rev*; pHXB2Bam-p3, ref. 13) and the expression vectors pcREV, pcREX, pc21, or pcTAX. Where only one plasmid is shown, the control plasmid pBC12/CMV/IL-2 was added to maintain a constant DNA concentration. Nine days after transfection, viral replication was assessed by measurement of the supernatant levels of reverse transcriptase³⁰ (a) and HIV-1 p24 gag protein³¹ (b). Similar results were obtained in two other experiments.

pc21 (lane 4) and pcTAX (lane 5). The pcTAX plasmid encodes the p21 protein as well as tax. These findings suggest that the p21 protein, which corresponds to a truncated form of the rex protein, could show a limited rex-like biological activity at high levels of expression.

As an independent and more rigorous test of whether the rex protein of HTLV-I could replace the *rev* function of HIV-1, we compared the capacity of the *rev* and *rex* expression vectors to rescue in *trans* the replication of a defective HIV-1 provirus specifically mutated in the *rev* gene¹³ (Fig. 3). Viral replication was monitored by the measurement of reverse transcriptase enzymatic activity and HIV-1 p24 gag-encoded protein in the culture supernatants. Transfection of an infectious wild-type clone of HIV-1 (pHXB2gpt)¹³ resulted in the release of high levels of reverse transcriptase and p24 protein into the supernatant as expected. But the supernatant concentrations of these proteins in cultures transfected with the *rev*-defective form of this same HIV-1 plasmid (HIV-1 Δ *rev*)¹³ did not rise above background. Cotransfection of either the pcREV or the pcREX expression vectors resulted in the rescue of viral replication of HIV-1 Δ *rev*, but neither the pc21 nor the pcTAX expression vectors restored replication of this defective provirus.

These studies provide unexpected evidence that the *rex* gene of HTLV-I can functionally replace the *rev*-encoded protein of HIV-1, leading to the effective production of the *env*, *gag* and *pol* gene products. The *rev* and *rex* gene products are both phosphoproteins^{19,20} primarily localized in the nuclei of expressing cells^{20,21}. Despite these general similarities, the *rex* and *rev* genes share no marked nucleotide or amino-acid homology. Furthermore, computer-generated predictions of secondary structure reveal no significant similarities. In evolutionary terms, HTLV-I and HIV-1 represent distinct viruses that are only distantly related²²⁻²⁶. On the basis of homology of the structural proteins encoded by members of the retrovirus family, HTLV-I and HIV-1 have been assigned to different virus sub-groups²⁶. Our findings demonstrate that a general strategy for the control of viral structural gene expression has been highly preserved in these viruses, despite marked evolutionary divergence in other functions, and possibly reflects involvement of a well conserved host factor in the function of both *rev* and *rex*. Functional similarities could alternatively be maintained by a requirement for direct recognition of a similar or identical viral target sequence by *rex* and *rev*. The conserved nature of *rex* and *rev* function may permit the future design of antiviral agents that

effectively inhibit the replication of both pathogenic retroviruses.

We thank Dr Richard Randall and Randy Fenrick for assistance with computer analysis of the *rev* and *rex* proteins and Drs Mark Feinberg and Flossie Wong-Staal for the HTLV-I derived pX cDNA used to prepare the pcREX, pcTAX and pc21 expression vectors and the pHXB2gpt plasmid containing an HIV-1 proviral isolate.

Received 18 July; accepted 9 September 1988.

1. Wong-Staal, F. & Gallo, R. C. *Nature* **309**, 640-642 (1985).
2. Sodroski, J. G. *et al. Science* **227**, 171-173 (1985).
3. Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* **229**, 69-73 (1985).
4. Sodroski, J. G., Rosen, C. A. & Haseltine, W. A. *Science* **225**, 381-385 (1984).
5. Fujisawa, J., Seiki, M., Kiyokawa, T. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2277-2281 (1985).
6. Cann, A. J., Rosenblatt, J. D., Wachsmann, W., Shah, N. P. & Chen, I. S. Y. *Nature* **318**, 571-574 (1985).
7. Felber, B. K., Paskalis, H., Kleinman-Ewing, C., Wong-Staal, F. & Pavlakis, G. N. *Science* **229**, 675-678 (1985).
8. Rosen, C. A. *et al. Nature* **319**, 555-559 (1986).
9. Cullen, B. R. *Cell* **46**, 973-982 (1986).
10. Chen, I. S. Y. *Cell* **47**, 1-2 (1986).
11. Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. *Nature* **330**, 489-493 (1987).
12. Sodroski, J. G. *et al. Nature* **321**, 412-417 (1986).
13. Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807-817 (1986).
14. Hidaka, M., Inoue, J., Yoshida, M. & Seiki, M. *EMBO J.* **7**, 519-523 (1988).
15. Knight, D. M., Flomerfelt, F. A. & Gharyeh, J. *Science* **236**, 837-840 (1987).
16. Rosenblatt, J. D. *et al. Science* **240**, 916-919 (1988).
17. Yoshida, M. & Seiki, M. *Rev. Immun.* **5**, 541-549 (1987).
18. Malim, M. H., Hauber, J., Fenrick, R. & Cullen, B. R. *Nature* **335**, 181-183 (1988).
19. Hauber, J., Bouvier, M., Malim, M. H. & Cullen, B. R. *J. Virol.* (in the press).
20. Kiyokawa, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8359-8363 (1985).
21. Cullen, B. R. *et al. J. Virol.* **62**, 2498-2501 (1988).
22. Wain-Hobson, S., Alizon, M. & Montagnier, L. *Nature* **313**, 743 (1985).
23. Chou, I.-M. *et al. Nature* **31**, 366-368 (1985).
24. Gonda, M. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4007-4011 (1986).
25. Sagata, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 677-681 (1985).
26. McClure, M. A., Johnson, M. S., Feng, D. F. & Doolittle, R. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2469-2473 (1988).
27. Gallo, R. C., Wong-Staal, F., Montagnier, L., Haseltine, W. A. & Yoshida, M. *Nature* **333**, 504 (1988).
28. Cullen, B. R. *Meth. Enzym.* **152**, 684-704 (1987).
29. Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
30. Poesz, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7415-7419 (1980).
31. Veronese, F. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5199-5203 (1985).

Stabilization of protein structure by interaction of α -helix dipole with a charged side chain

Daša Šali, Mark Bycroft & Alan R. Fersht*

Department of Chemistry, Imperial College of Science and Technology, London SW7 2AY, UK

The α -helix in proteins has a dipole moment resulting from the alignment of dipoles of the peptide bond which can perturb the pK_a s of ionizing groups. One of the two histidine residues (His18) in barnase, the small ribonuclease from *Bacillus amyloliquefaciens*, is located at the negatively charged end (C-terminal) of an α -helix. From NMR titrations of wild-type and engineered mutants we find that the pK_a of His18 is 7.9 in wild-type enzyme, 1.6 units above the value in the urea-denatured enzyme and in model peptides. This implies that there is a favourable interaction between the protonated form of His18 and the α -helix that should stabilize the native structure at neutral pH by 2.1 kcal mol⁻¹. Denaturation at various values of pH of wild-type and mutant enzymes engineered at position 18 shows that this is so. The increase in stability of the enzyme as the pH changes from 8.5 to 6.3 is attributable to this interaction, and the pH-stability curve fits pK_a values for His18 in native and urea-denatured enzymes that are consistent with the NMR data.

* To whom correspondence should be addressed. Present address: University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK.

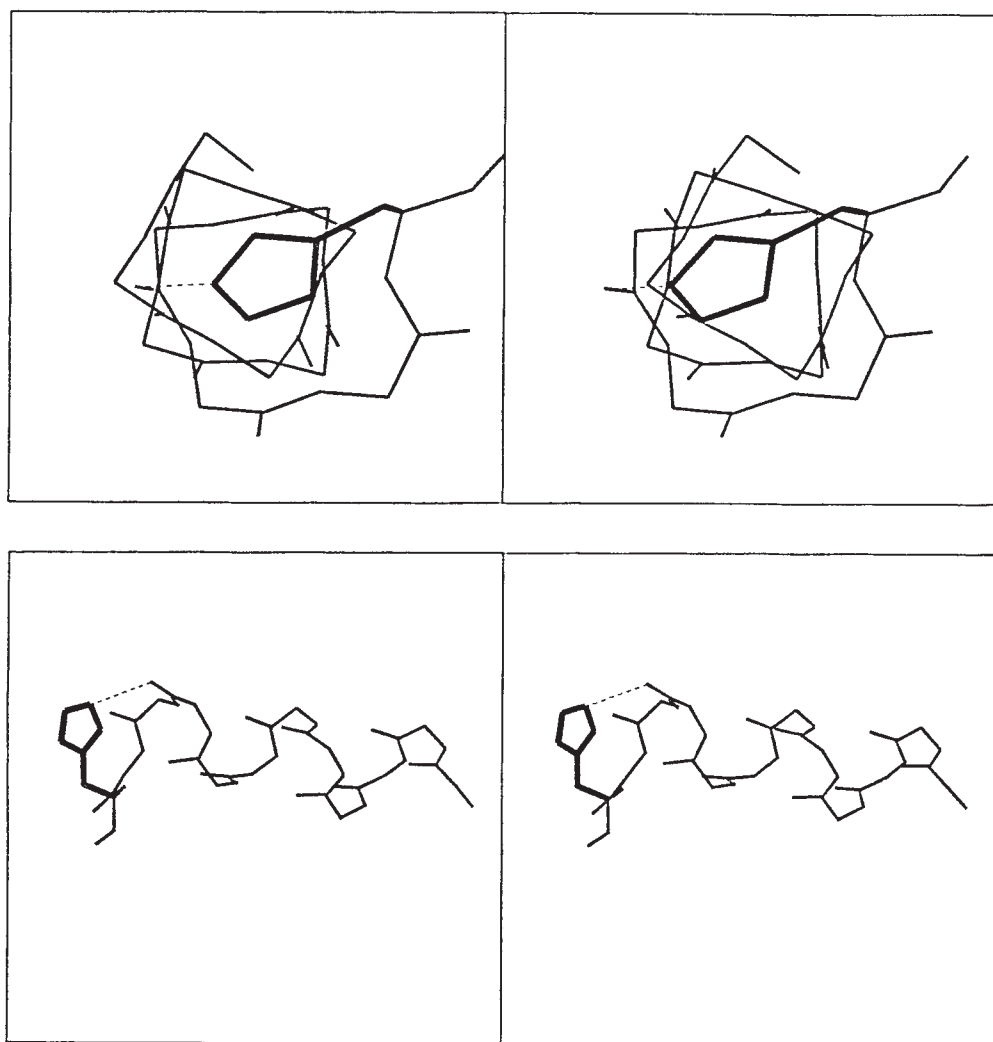


Fig. 1 Two stereo-pairs His18 and its relationship with the α -helix. The hydrogen bond between $N_\epsilon(H)$ of His18 and the backbone carbonyl oxygen of Gln15 is indicated.

The stabilization energy of proteins is relatively small; the observed values of some 5–20 kcal mol⁻¹ represent the differences between two large numbers—the free energies of the folded and unfolded states—which are closely balanced, so whether or not a protein folds depends crucially on the presence of small interaction energies. Different interactions have been proposed to stabilize enzymes, but accurate experimental data are required to quantify their contributions. We have used barnase as a paradigm for experimental studies on protein stability and folding because of its favourable properties¹. The enzyme consists of a single polypeptide chain of 110 residues with no disulphide crosslinks and undergoes reversible unfolding². We have used protein engineering to remove (or add) defined interactions that contribute to the stability of the native enzyme, and have then measured the change in free energy of folding from urea-induced denaturation. This provides an empirical measure of changes in stability on change of structure and a data base for refining theoretical calculations.

Long-range electrostatic interactions are important in both catalysis and protein stability, but the magnitudes of charge-charge interactions were until recently difficult to calculate because of a lack of experimental data. Protein engineering has been used to mutate charged surface residues of subtilisin³. Another important electrostatic element could be the dipole that results from the aligned peptide units of the α -helix^{4–6}: this can participate in electrostatic interactions which affect the binding of charged ligands^{6,7}, the pK_a s of ionizing groups⁸ and the stability of the helix itself^{9–12}. An anomalously high pK_a of a

histidine residue in haemoglobin has been measured by NMR and the perturbation attributed to the field of an α -helix⁸. Negatively charged residues tend to cluster at the positively-charged end of an α -helix and vice versa^{9,10}. Changes in T_m of small model helices show qualitatively that the helices are stabilized by dipole-charge interactions^{11,12}. There is compelling evidence for the importance of electrostatic effects from helix dipoles, but the precise magnitudes are unknown⁶. Here we provide quantitative measurements of the energy of a charge-dipole interaction that stabilizes barnase.

There are two histidine residues in barnase, His18 and His102. His18 is located at the C-terminal end of a helix that extends from residues 6 to 18 (ref. 13) (Fig. 1). The two ring nitrogen atoms of His18 are between 3 and 5.5 Å from the backbone C, N and O atoms of residues 14, 15, 16 and 17 at the end of the α -helix, and the H attached to N_ϵ makes a hydrogen bond with the backbone carbonyl oxygen of Gln 15. As is usual in proteins, the helix is partially shielded from water, unlike simple model helices. NMR titration (Fig. 2) shows that one histidine ionizes at a nearly normal pK_a of 6.30 (± 0.05), whereas the other has a perturbed pK_a of 7.86 (± 0.05). By analogy with the experiment on haemoglobin⁸, we expect the perturbed residue to be His18. We have investigated the possible stabilization of the protein by the helix dipole-charge interaction by measuring the pK_a of each histidine in both native and urea-denatured enzyme by NMR and assigning them using mutants lacking each of the histidine residues. The importance of the electrostatic interaction between the positive charge on the protonated imidazole of

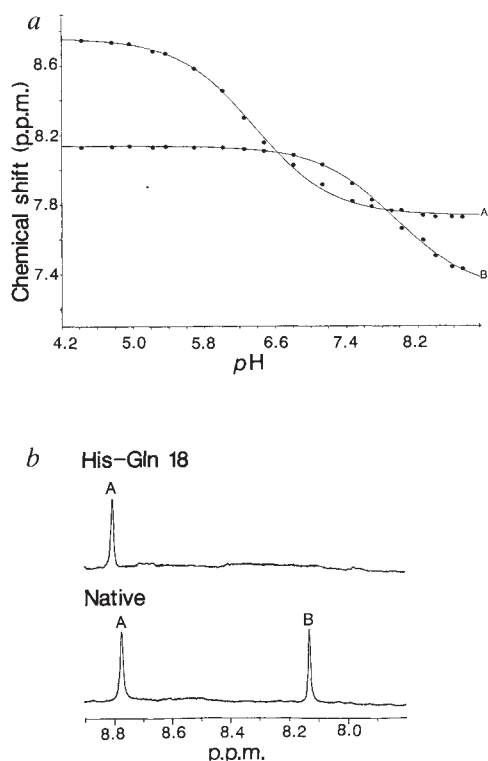


Fig. 2 *a*, Titration curves of C2-H resonances of histidine residues 18 (B) and 102 (A) in native barnase. *b*, Spectra (500 MHz ^1H -NMR) of histidine C2-H resonances of native barnase at pH 4.5 and mutant His18 to Gln18 at pH 4.3. NMR titrations were performed on 1 mM protein using a Bruker AM500 spectrometer at $25 \pm 1^\circ\text{C}$ in D_2O , using NaOD and DCl to adjust the pD, with no added salt. Titration of imidazole in D_2O and H_2O showed that the measured pK_a is higher in D_2O by 0.07 units when the pH meter/glass electrode is calibrated with standard buffers in H_2O . The calculated value for the pK_a of His18 in D_2O is 7.93 ± 0.05 . This is equivalent to 7.86 ± 0.05 in H_2O . Urea was found to affect the titration of acids and bases in a parallel manner. Standard pH 7.00 phosphate reads 7.30 ± 0.02 in 5 M urea and the pK_a of imidazole increases by 0.29 ± 0.02 units in 5 M urea when calibrated against standard buffers in H_2O (Tris and MES (2-(*N*-morpholino) ethanesulphonic acid) also change their pK_s in an identical manner in urea, see legend to Fig. 4). The measured value of the pK_a of His18 in 5 M urea calibrated against standard buffer in H_2O is 6.59 ± 0.05 . This is equivalent to 6.30 ± 0.05 in H_2O . Wild-type and mutant enzyme were prepared as before¹. The following primer was used to direct the mutation His to Gln18: 5'-GTAGCTTT*TGATATGTC-3' (where the asterisk follows the mismatch).

His18 and the enzyme was measured by determining the changes in free energy of folding as the imidazole titrates with pH.

Assignment of pK_a s and interaction of His18 with enzyme. A mutant having Gln substituted for His at position 18 was constructed and titrated, when the histidine of pK_a 7.86 disappeared and so the high pK_a was assigned to His18. Titration of the proteins in 5 M urea shows that the pK_a of His18 drops to 6.30 ± 0.05 , the normal value in small peptides⁸, so the pK_a of His18 is, therefore, perturbed by 1.6 units by interaction with the protein. This implies that the protonated form of the imidazole is stabilized by the protein by $2.1 (\pm 0.1)$ kcal mol⁻¹ relative to the uncharged form.

We infer from the simple thermodynamic cycle in Fig. 3 that the protein is stabilized by 2.1 kcal mol⁻¹ on protonation of His18 as follows. If the equilibrium constant for unfolding of the enzyme containing the protonated form of His18 is K_{unfold} ,

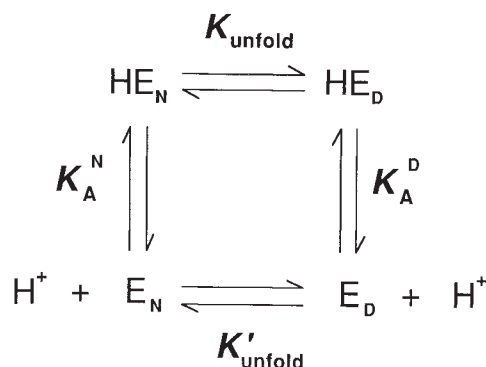


Fig. 3 Thermodynamic cycle relating the acid dissociation constants in the native and denatured forms of the enzyme, K_A^N and K_A^D , with the equilibrium constants for unfolding of the enzyme containing the protonated and unprotonated forms of a residue, K_{unfold} and K'_{unfold} , respectively.

for the unprotonated form is K'_{unfold} , and the ionization constant for His18 in the native enzyme is K_A^N and in the denatured form is K_A^D then

$$K_{\text{unfold}}/K'_{\text{unfold}} = K_A^N/K_A^D \quad (1)$$

That is, the equilibrium constant for unfolding changes in parallel with the difference in ionization constants in the native and denatured states. As the pK_a of His102 is 6.3 in both the native and urea-denatured forms, this ionization does not affect the unfolding equilibrium.

Destabilizing the protein by mutation of His18 to Gln18. A crude way of measuring the interaction energy between the protonated His18 and the enzyme is to compare its stability with a mutant in which His18 is substituted by a neutral side chain. But mutation of His18 to Ala18 loses not only the electrostatic interaction, but also the hydrogen bond from the $N_\epsilon(\text{H})$ of the imidazole ring to a main-chain carbonyl oxygen, as well as hydrophobic contacts. The mutant His18 to Ala18 is so destabilized that we were unable to isolate enough for analysis, which we have found before for other mutants where the destabilization is >4 – 5 kcal mol⁻¹. Mutation of His18 to Gln18 is more appropriate because as shown previously in the tyrosyl-tRNA synthetase, its carboxamide $-\text{NH}_2$ may form hydrogen bonds equivalent to those of the N_ϵ of histidine¹⁴ and can also make hydrophobic contacts. Wild-type enzyme is more stable than the His18 to Gln18 mutant at pH 6.3 by 1.6 kcal mol⁻¹. However, it could be argued that the similarity to the expected value for the electrostatic interaction is fortuitous: the dipole from the $-\text{CONH}_2$ group could well make a dipole/dipole interaction with the helix, but this is counterbalanced by the loss of internal entropy of rotation in the flexible glutamine side chain on hydrogen bonding back to the main chain. But the mutant His18 to Gln18 provides a control for the following improved procedure.

Variation of K_{unfold} with pH gives electrostatic stabilization energy. The most precise way of using denaturation to determine the interaction energy is to measure the change in stability as the histidine titrates with pH. As seen in Fig. 4*a*, addition of urea to wild-type enzyme generates a family of parallel denaturation curves as the enzyme is titrated from pH 6.3 to 8.5, whereas the stability of the mutant His18 to Gln18 varies only slightly. A plot of K_{unfold} against pH (Fig. 4*b*) for wild-type enzyme fits an ionization curve of pK_a 7.94 ± 0.14 . The ratio of $K'_{\text{unfold}}/K_{\text{unfold}}$, as defined in Fig. 3, is 16 ± 4 . Substituting these values into equation (1) gives a value of 6.7 ± 0.2 for the pK_a of His18 in the denatured enzyme. There is a small but systematic

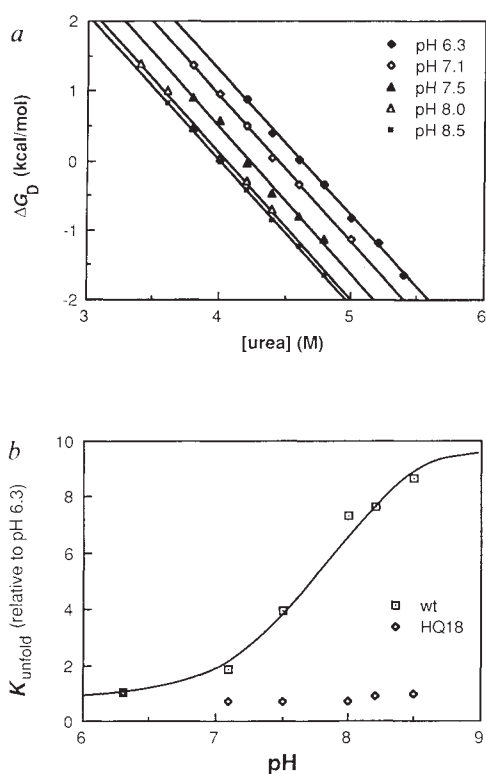


Fig. 4 *a*, Free energy of denaturation of wild-type barnase at different values of pH plotted against urea concentration according to the standard equation: $\Delta G_D = \Delta G_D^{\text{H}_2\text{O}} - m[\text{urea}]$. The data at pH 8.2 are omitted because of overcrowding the graph. *b*, Plot of the equilibrium constant for unfolding, K_{unfold} , as a function of pH for wild-type enzyme and the mutant His to Gln18. The observed values of K_{unfold} are given relative to the value at pH 6.3 for both enzymes. (Note that His18 to Gln18 is less stable than wild type by $1.6 \text{ kcal mol}^{-1}$ at pH 6.3, but by only $0.3 \text{ kcal mol}^{-1}$ at pH 8.5.) The data for wild-type enzyme fit to an ionization curve of $pK_a 7.94 \pm 0.14$ and a ratio of $K'_{\text{unfold}}/K_{\text{unfold}}$ (as defined in Fig. 3) of 16 ± 4 . Denaturation was observed spectrofluorimetrically with excitation at 290 nm and emission at 315 nm. The fluorescence falls by some 70% on denaturation. Experiments were performed at 25 °C in 50 mM MES acid/base buffer at pH 6.3 and Tris-HCl at higher values of pH (reducing the concentration of buffer tenfold does not affect the results). The concentration of barnase was between $5\text{--}10 \text{ } \mu\text{g ml}^{-1}$. The change in free energy of folding as a function of pH, $\Delta\Delta G_D$, is calculated very simply and accurately because the denaturation curves overlap. In a manner analogous to that described previously¹, $\Delta\Delta G_D$ between two pH values pH_x and pH_y , $\Delta\Delta G_D(pH_x - pH_y)$, is calculated from the direct measurement of the fractions of unfolded and folded states of the protein at identical concentrations of urea:

$$\Delta\Delta G_D(pH_x - pH_y) =$$

$$RT \ln \{ ([\text{folded}]/[\text{unfolded}]pH_y) / ([\text{folded}]/[\text{unfolded}]pH_x) \}$$

As there is just a small change in stability from one value of pH to the next, the ratios of $[\text{folded}]/[\text{unfolded}]$ can readily be measured over a range of values of urea concentrations. The values of the pK_a s of phosphate (the standard calibration buffer), Tris and MES all increase in 4 M urea by 0.23 ± 0.02 units when calibrated against standard buffers in H_2O . Thus, when calibrated against standard buffers in urea, measured values of pH and pK_a s are unaffected by the addition of urea. To a good approximation, therefore, urea does not perturb the ionization of groups in the protein. Even if there are small effects (± 0.02 units at 4 M urea) on the different types of acids and bases, these are expected to be the same in both native and denatured states and so would not perturb the denaturation equilibrium.

variation of the observed unfolding constant for the mutant His18 to Gln18, with pH ($\pm 20\%$) that may reflect other ionizations in the protein or environmental effects. To normalize for these in the wild-type enzyme, the ratio of $(K_{\text{unfold}})_{\text{wt}}/K_{\text{unfold}}^{\text{His}\rightarrow\text{Gln18}}$ should be plotted. This gives a slightly poorer fit (not shown), with values of pK_a of 7.7 ± 0.2 for His18 in native enzyme and 6.35 ± 0.25 in the denatured enzyme. The value of $K'_{\text{unfold}}/K_{\text{unfold}}$ is 22 ± 11 . The change in interaction energy of His18 with the protein on ionization is $1.4\text{--}2.1 \text{ kcal mol}^{-1}$ (that is, it equals $RT \ln (K'_{\text{unfold}}/K_{\text{unfold}})$). These values are consistent with the NMR data within experimental error.

Data are internally consistent. The two sets of experiments measure the two different pairs of thermodynamic constants in Fig. 3. NMR measures K_a^N and K_a^D whereas denaturation measures K'_{unfold} and K_{unfold} . In addition, the pH dependence of denaturation gives K_a^N and K_a^D . The two sets of experiments thus provide a test of the applicability of the simple two-state scheme of Fig. 3 to the denaturation of barnase and the suitability of measuring denaturation. The NMR data are inherently the more accurate, however, and so the thermodynamic constants derived from those should preferably be used.

The electrostatic interaction energy of His18 is predominantly with the α -helix. We have been careful to state so far that the electrostatic interaction energy is between His18 and the (whole) protein because long-range interactions must be made with all the charged groups. But, it does appear that the energy can be attributed predominantly, and probably solely, to the electrostatic interaction of His18 with the peptide dipoles that constitute the α -helix. Firstly it can be seen that positively and negatively charged side chains in the protein are not close to His18 and are reasonably equally distributed from it: the closest charged groups are the ϵ - NH_2 of Lys19, 9 Å away, and the carboxylate of Asp12, 12 Å away. Indeed, preliminary calculations (S. Wodak, personal communication) indicate that the net field from the charged residues of the protein is close to zero at His18. Secondly, there is direct evidence from the effects of counterions on the pK_a of His18; it is well known from classical measurements that counterions mask electrostatic effects from charges in proteins. Further, recent accurate measurements of effects of mutation of charged groups on the surface of subtilisin show that 1 M KCl almost exactly screens the charge on individual side chains (see ref. 3 for example). The pK_a of His18 increases from 7.86 in 1 mM salt to 7.98 in 1 M KCl, a small increase similar to that found for the titration of imidazole at low and high salt. Thus the high pK_a of His18 does not result from interactions with charged side chains. The lack of dependence of the electrostatic interaction on salt suggests that the field of the helix dipole is only weakly shielded by counterions.

We thank Dr R. W. Hartley for his gift of the barnase expression system. This work was supported by the MRC and EEC (B.A.P.).

Received 8 August; accepted 12 September 1988.

1. Kellis, J. T. Jr, Nyberg, K., Sali, D. & Fersht, A. R. *Nature* **333**, 784–786 (1988).
2. Hartley, R. W. *Biochemistry* **7**, 2401–2408 (1968); *Biochemistry* **8**, 2929–2932 (1969).
3. Thomas, P. G., Russell, A. J. & Fersht, A. R. *Nature* **318**, 375–376 (1985); Russell, A. J. & Fersht, A. R. *Nature* **328**, 496–500 (1987).
4. Wada, A. *Adv. Biophys.* **9**, 1–63 (1976).
5. Hol, W. G. H., Van Duijnen, P. T. & Berendsen, H. J. C. *Nature* **273**, 443–446 (1978).
6. Hol, W. G. J. *Prog. Biophys. molec. Biol.* **45**, 149–195 (1985).
7. Warwicker, J. & Watson, H. C. *J. molec. Biol.* **157**, 671–679 (1982).
8. Perutz, M. F., Gronenborn, A. M., Clore, G. M., Fogg, J. H. & Shih, D. T.-b. *J. molec. Biol.* **183**, 491–498 (1985).
9. Blagdon, D. E. & Goodman, M. *Biopolymers* **14**, 241–245 (1975).
10. Wada, S. & Nakamura, H. *Nature* **291**, 757–758 (1981).
11. Shoemaker, K., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2349–2353 (1985); *Nature* **326**, 563–567 (1987).
12. Mitchinson, C. & Baldwin, R. L. *Proteins* **1**, 23–33 (1986).
13. Mauguen, Y. *et al. Nature* **297**, 162–164 (1982).
14. Lowe, D. M., Fersht, A. R., Wilkinson, A. J., Carter, P. & Winter, G. *Biochemistry* **24**, 1506–1509 (1985).

Corrigenda

Thyroid hormone receptor α isoforms generated by alternative splicing differentially activate myosin HC gene transcription

Seigo Izumo & Vijak Mahdavi

Nature 334, 539–542 (1988).

IN the typing of the derived amino-acid sequence of rTR α 1 and rTR α 2 cDNAs, the sequence for amino acids 401–410 were inadvertently swapped between the two isoforms. The correct sequence reads: FLEVFEDQEV for rTR α 1 and QLLGMHVQVQ for rTR α 2.

T-cell antigen receptor genes and T-cell recognition

Mark M. Davis & Pamela J. Bjorkman

Nature 334, 395–402 (1988).

IN this Review Article, the colour-coding given in the legend to Fig. 4 relates to an earlier version of the figure, not the one in the final version of the manuscript. The legend is reprinted here in full with the correct colour coding:

Fig. 4 Representations of the structures of *a*, the immunoglobulin combining site; *b*, the peptide-binding site of an MHC molecule; and *c*, the alignment of CDRs in a hypothetical TCR over a peptide-MHC complex. *a*, The arrangement of CDRs in an immunoglobulin antigen-binding site (Fab J539)¹²² viewed from above (from the position of the antigen). The carbon- α backbone of V_H and V_L is shown in blue with van der Waals' surfaces highlighting the three CDRs from each domain (CDR1: blue, CDR2: yellow, CDR3: pink). The first and second CDRs from one variable domain (for example, immunoglobulin V_L or TCR V_α or V_γ) are separate from their counterparts on the partner variable domain (immunoglobulin V_H or TCR V_β or V_δ). The space between them is occupied by the third CDR from each V domain. The similarities between immunoglobulins and TCRs suggest that TCRs may have a combining site that preserves these same general features (see text for details). *b*, Top surface of an MHC molecule with a (hypothetical) bound peptide. The carbon- α backbone of the α_1 and α_2 domains of HLA-A2 (ref. 40) is shown in blue with van der Waals' surfaces highlighting the two α -helices (yellow). Van der Waals' surface of a hypothetical bound peptide (a 12-mer polyvaline α -helix) that has been fitted between the HLA α -helices is shown in pink. The putative recognition site for processed foreign antigens is located on the top surface of the molecule between two α -helices for the human class I histocompatibility molecule HLA-A2^{40,41}. The N-terminal α_1 and β_1 domains of class II MHC molecules are predicted to have a similar tertiary structure and peptide-binding site⁴². Note that the distance between the MHC α -helices, and the separation of the first and second CDRs of each V -region in the combining site shown in part *a* are very similar. (Figures are to scale with respect to each other). *c*, Model for TCR interaction with a peptide-MHC complex. The combining site CDRs (*a*) are shown aligned over the peptide-MHC complex (*b*). The molecules in this figure are rotated $\sim 90^\circ$ with respect to their orientations in (*a*) and (*b*). V domains of Fab J539 (ref. 122; top of figure) here represent the V domains of a TCR bound to a peptide-MHC complex (bottom of figure). The carbon- α backbone of V_L and V_H (blue) is shown with main and side-chain atoms of CDR1 and CDR2 (yellow) from each domain. The main and side-chain atoms of the CDR3s of each domain are shown in pink. The carbon- α backbone of the α_1 and α_2 domains of HLA-A2 (blue) is shown with the main and side-chain atoms of the two α -helices in yellow and the main and side-chain atoms of a hypothetical peptide bound between the two helices in pink. The relatively flat surface of the V -region combining site is complementary to the flat surface of the MHC-peptide complex, with CDR1 and CDR2 of each V domain 'fitting' over an MHC α -helix. The CDR3s from each V domain are then aligned over the peptide (see text for details). Because the antibody (and presumably TCR) V regions pair with approximate twofold (dyad) symmetry^{9,10}, a similar interaction of CDR1 and CDR2 with the MHC α -helices and CDR3 with peptide would be accomplished if the V_H - V_L dimer shown in this figure were rotated by 180° about the pseudo-dyad axis between the two V regions (axis is vertical in this figure). Note that the relative sizes of the combining site (*a*) and the peptide-binding site of the MHC molecule (*b*) would allow TCRs to be bound in different registers along the MHC α -helices, depending on the particular peptide bound to that MHC protein. Thus the V genes employed by different T cells restricted to the same MHC molecule would not be expected to be identical. Furthermore, the C-terminal few amino acids encoded by a V gene are part of the V -J or V -D-J junction, and would be predicted by this model to interact primarily with peptide rather than MHC determinants.

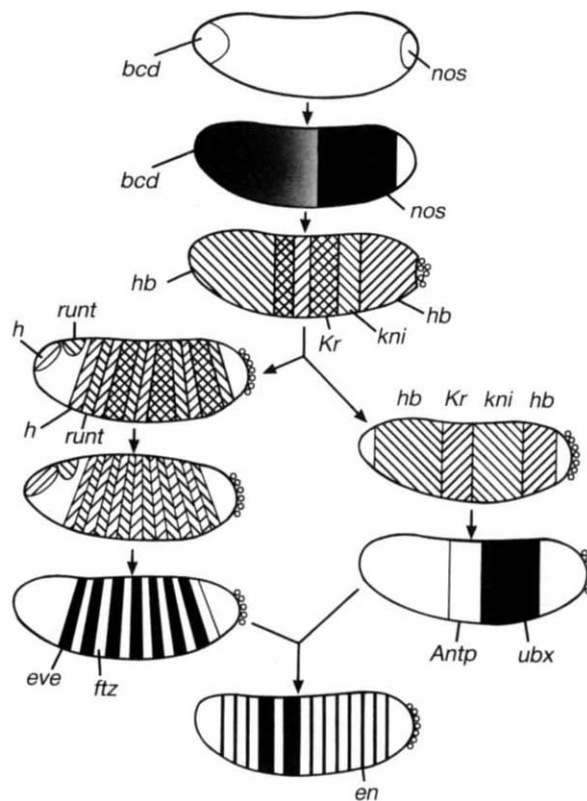
In addition, in Table 1 on page 397, the penultimate column should be headed γ rather than ν .

The molecular genetics of embryonic pattern formation in *Drosophila*

P. W. Ingham

Nature 335, 25–34 (1988).

THE labels were omitted from Fig. 2 in this Review Article, and are shown below in a black-and-white version of the figure.



Errata

Sensory transmitters regulate intracellular calcium in dorsal horn neurons

M. D. Womack, A. B. MacDermott & T. M. Jessell

Nature 334, 351–353 (1988).

IN this letter, the two sentences beginning on line 29 on page 353 should read: "Both classes of transmitters appear to regulate $[Ca^{2+}]_i$ in these neurons, albeit by many different mechanisms. Substance P increases $[Ca^{2+}]_i$ in many dorsal neurons by mobilizing intracellular Ca^{2+} stores."

Global fire at the Cretaceous–Tertiary boundary

W. S. Wolbach, I. Gilmour, E. Anders, C. J. Orth & R. R. Brooks

Nature 334, 665–669 (1988).

ON page 666, the third line from the bottom should read: "The soot content rises even more sharply across the boundary: 11, 1930 and 97 p.p.m."

A protein sequence/structure database

*Protein Engineering Club Database Group**

Researchers have been wishing for years for a database integrating protein sequences and protein structures. A group from the United Kingdom has achieved this long-sought objective.

COHERENT handling of the explosion in the amount of protein sequence and structural data has become a major problem for molecular biologists¹. The value of protein databases is well established: sequence databases enable the detection of homologies, and structural databases aid in the understanding of protein structure and modelling. But there is great potential for the integration and expansion of these databases into a fuller, more accessible knowledge resource² which could be used to develop theories of protein structure, to design protein engineering experiments, and to enable the function of genes to be deduced from their nucleotide sequences. Much more information about a particular protein can be gleaned when sequence and structure data are combined, and when the relationships between them are systematically identified and incorporated.

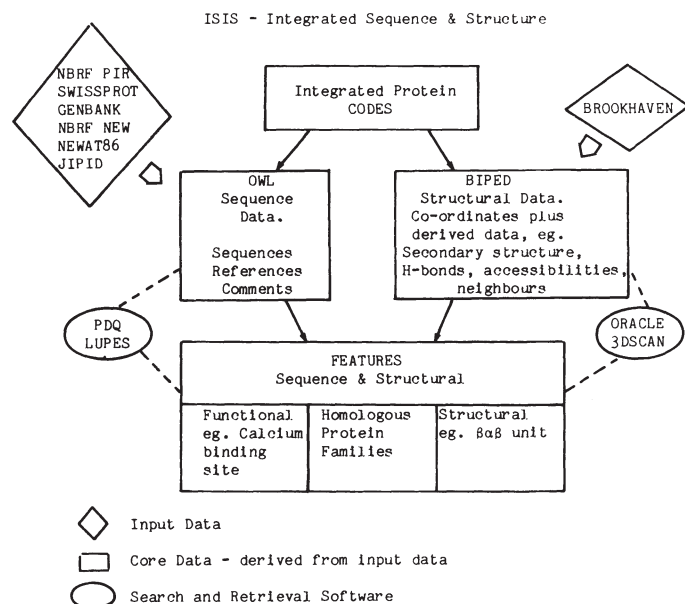
Several groups worldwide are currently involved in developing structural databases, including Wodak at the Université Libre de Bruxelles, Belgium, Hirayama and Ceska of Kyowa Hakko Kogyo Co. Ltd, Tokyo, Japan, and Go and collaborators at Kyoto University, Japan. At Leeds and Birkbeck we have developed an integrated data resource of protein sequence and structure, named Integrated Sequences/Integrated Structures (ISIS), which we hope will be the first of many protein knowledge resources to come. ISIS not only incorporates the basic sequence and coordinate data, organized into a rational manner without redundancies, but also includes derived data on protein properties and relationships. ISIS employs appropriate codes and formats to enable different data sets to be cross-referenced, and is accessed by flexible search and interrogation software.

The first release of ISIS is now available. The project has been supported by the UK Protein Engineering Club, established to support basic research by the Science and Engineering Research Council (SERC) and four companies: Celltech, Glaxo, ICI and Sturge/RTZ.

Integrated system

The components of the system (Fig. 1) are core sequence and structural data, derived

Fig. 1 The ISIS integrated protein sequence and structure data resource. Source data: **NBRF PIR** and **NEW 16-0**: National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road, N.W., Washington, DC 20007, USA⁴. **SWISS-PROT 6-0**: A. Bairoch, Département de Biochimie Médicale, Centre Médicale Universitaire, 1211 Geneva 4, Switzerland. **GenBank 54**: Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA (translations used in part programs of J.W. Fickett, 1986)⁵. **NEWAT 86**: R.F. Doolittle, Department of Chemistry, University of California San Diego, La Jolla, California 92093, USA. **JIPID**: Japanese International Protein Information Database, Science University of Tokyo, Japan. **BROOKHAVEN October 1987**: Brookhaven National Laboratory, Upton, New York 11973, USA.



data including features, and search and retrieval software. The composite sequence database OWL is an integration of the publicly available databases, together with a translation of the GenBank nucleic acid sequence library. The source databases have been merged, and redundant and trivially different entries have been eliminated according to defined criteria by using a suite of management programs called COMPO. The better-validated and annotated source databases have been given higher priority. OWL is in the National Biomedical Research Foundation (NBRF) format for compatibility with established search and retrieval software such as PSQ and FASTP. The NBRF format is also compatible with the PDQ rapid information retrieval software and a package called LUPES which uses pattern discriminators for knowledge-based database exploration. OWL is updated at three-month intervals, following each release of the NBRF source database.

The BIPED relational database of 3-dimensional structure is based on protein coordinates and crystallographic data from the Brookhaven Protein Data Bank³. The coordinates are processed to remove trivial errors, to highlight chain discontinuities and to indicate missing atoms and residues. Powerful additional data derived from the coordinates, such as secondary structures, torsion angles, solvent accessibilities, hydrogen bonds, salt

bridges, disulphide bridges and nearest neighbours are also included. Data are organized into tables reflecting the hierarchical organization of protein structure, with information at the protein, chain, residue and atom levels. The Oracle Relational Database Management System (Oracle Corporation UK Ltd, Bristol, UK) has been used to store the tables.

Enquiries use the simple commands of the database language SQL. Information from one or more proteins can be extracted and data in one or more tables can be cross-correlated to meet a wide range of queries. (For example, it took only 14 seconds of MicroVAX II (Digital Equipment Company Ltd, Reading, UK) CPU time to extract the Φ and Ψ angles of all prolines in the fourth positions of alpha-helices for 296 proteins.) Information in the tables is also available as flat files to allow incorporation into non-relational database systems.

We are currently developing an independent non-relational SQL-based database system called 3DSCAN in order to facilitate the extraction of structural features. This will be developed for a microcomputer. Integration of the OWL, BIPED and 3DSCAN databases will be achieved through a system of protein codes and identifiers.

Features

The data resource incorporates features of

*D. Akrigg, A.J. Bleasby, N.I.M. Dix, J.B.C. Findlay, A.C.T. North, D. Parry-Smith, J.C. Wootton, The University, Leeds LS2 9JT, UK. T.L. Blundell, S.P. Gardner, F. Hayes, S. Islam, M.J.E. Sternberg, J.M. Thornton, I.J. Tickle, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX. P. Murray-Rust, Glaxo Group Research, Greenford, Middlesex UB6 0HE, UK.

protein sequence and structure. The term "feature" is deliberately unspecific and refers to any structure or substructure of proteins that can be identified by a set of aligned sequences or superposed 3-dimensional structures. ISIS will include tables of structural features and a library of sequence features by the second release. Some structural features are commonly called "motifs", "supersecondary structures" or "domains". Some have characteristic interactions, for example, nucleotide binding pockets, DNA binding structures and glycosylation sites. Many of the features lack 3-dimensional structural information, and have been assigned from biochemical and mutational evidence, the alignment of a family of sequences, and conserved residue patterns. Probably many more are implicit in the sequence database and remain to be found.

Sequence features are substructures of proteins defined by sequence alignments and by pattern recognition discriminators. A "discriminator" is a component of a pattern recognition system that discriminates (in most cases, imperfectly) a particular feature from all other protein sequences, and produces a "hit list" in which the known and predicted examples of a feature are of high rank. The simplest types of discriminator are a consensus sequence with defined redundancies, and a numerical weight matrix based on residue frequencies at each position of the known canonical examples of the feature.

LUPES software can generate and use these simple discriminators — and extensions of them in which additional positive and negative weights are added — to explore the database. The results obtained using different discriminators can be combined by logical operations. The "diagnostic power" of discriminators can be measured objectively from the results of database searches. LUPES also includes a number of programs for handling multiple sequence alignments.

Structural features are defined by superpositions in three dimensions and by sequence alignments. There are three major types of component tables: homologous protein families of known structure, based on structural alignments; structural features such as beta-strands and beta-hairpins; and functional features such as the calcium-binding EF hand structures.

Computer requirements

ISIS has been developed for the DEC VAX/VMS environment, but is structured for future adaptation to other hardware. It is available on tape, complete or in separate modules, at minimal cost to academic users and for a subscription to commercial companies. It is also being set up at the SERC Daresbury Laboratory in the United Kingdom for access via the SEQNET/JANET network in the near future. Three modules are available in the

Computing science

Over 90 companies will be represented at the Scientific Computing and Automation Conference next week in Philadelphia, Pennsylvania, exhibiting everything from hardware boards to a program for building a CD-ROM database.

SCIENTIFIC Solutions has a **data acquisition/process control board** for upgrading IBM PS/2 computers to the high performance digital and analogue I/O required for laboratory applications (*Reader Service No. 101*). The \$995 (US) model MC-DAS 1612 is a single-slot board that implements the PS/MicroChannel interface, performing burst mode DMA transfers as an I/O slave. Scientific Solutions says the board has a throughput of 80,000 samples per second in single channel and auto-scan modes, and 50,000 samples per second in normal mode, where non-contiguous channels are being converted. The MC-DAS 1612 board comes with an interface cable, screw terminal board and software which includes a menu-driven interface, demonstration programs, a dedicated DMA disk utility and the LabPac software support system, a memory resident software driver that allows applications to be developed in any language based on MS-DOS.

Molecular modelling

Jandel Scientific has a new software package for **modelling serial sections** which can construct whole graphic models from the sections and display the results

first release: the OWL sequence database (55 Mbytes); PDQ and LUPES sequence scan, pattern recognition and retrieval software (9 Mbytes); and BIPED, available as Oracletables (274 Mbytes) or as flat files (190 Mbytes). The Oracle Relational Database Management System (43 Mbytes) must be purchased from Oracle Corporation UK Ltd. The 3DSCAN software, additional features tables, and updated OWL, BIPED and sequence features will be available in the second release planned for early 1989. □

For more information, fill in reader service number 100, or contact Ian Robertson, British Technology Group, Electronics and Information Technology Division, 101 Newington Causeway, London SE1 6BU, UK. Scientific enquiries should be addressed to J. C. Wootton at Leeds for sequences, and J. M. Thornton at Birkbeck for structures. Network enquiries should be sent via electronic mail to SEQMAIL@UK.AC.DARESBUY.

1. Pabo, C.C. *Nature* **327**, 467 (1987).
2. Ponger, S. *Nature* **332**, 24 (1988).
3. Bernstein, F.C. *et al.* *J. molec. Biol.* **122**, 535 (1977).
4. George, D.G. *et al.* *Nucleic Acids Res.* **14**, 11 (1986).
5. Fickett, J.W. *Trends biochem. Sci.* **11**, 190 (1986).



Jandel's PC3D software models serial section for display and manipulation.

on the screen of a PC (*Reader Service No. 102*). Objects can be rotated, viewed from any angle or orientation, and "dissected" onscreen with Jandel's PC3D software, making it useful for applications such as untangling intertwined neurons. Serial sections of up to 400 polygons are entered from a digitizing tablet or supplied as ASCII data. Up to 1,000 sections can be combined into a single image, and PC3D can compensate for differing section increments and missing sections by repeating another section. Once entered, any combination of the components of the object may be displayed, in a unique solid or transparent colour. The software can automatically calculate the volume of each component, and an optional "movie" module is available to simulate the movement of the reconstructed sections. The PC3D software costs \$995 (US), and requires an IBM PC/XT/AT or compatible, 640K RAM, 8087 or 80287 math co-processor chip, IBM EGA or Tecmar Graphics Master graphics card, and a digitizing tablet.

Polygen has just released version 2.1 of QUANTA, its software for viewing and manipulating **3-dimensional molecular models** in real time (*Reader Service No. 103*). The new version has added capabilities in its conformational search and protein modelling subsystems. The conformational search subsystem generates large numbers of alternate conformations of molecules, which can be examined through energy minimization, molecular dynamics, graphical modelling, and distance constraints to see which structures are most likely to be important in the molecule's chemical or physiological activity under differing temperatures and surroundings. The search techniques used

to generate the conformations include systematic grid search, random sampling methods, and Monte Carlo methods, operating on all or a part of small and large molecules. The protein modelling enhancements implement the technique known as homology model-building, in which a protein of unknown 3-dimensional structure is built up, piece by piece, by matching its fragments to similar portions of proteins whose structures are known. QUANTA runs on the IRIS 4D- and 3000-series superworkstations from Silicon Graphics, and the GS1000 graphics supercomputer from Stellar Computer.

Software

LaserRETRIEVE is the name of a software package offered by Hewlett-Packard for **developing CD-ROM applications** (Reader Service No. 104). LaserRETRIEVE consists of a user-interface



Hewlett-Packard's software for developing CD-ROMs.

software module that retrieves information from a CD-ROM, and a database-building software module that indexes and structures a large amount of information from multiple sources into one database ready for CD-ROM mastering. Once the CD-ROM is mastered and replicated, LaserRETRIEVE can be used to search and retrieve specified information from the finished disk. CD-ROMs developed using LaserRETRIEVE can be searched by browsing through a general table of contents or by keywords. The program can handle both full-text and graphic documents. The database-building module of LaserRETRIEVE requires an HP Vectra PC or an IBM PC/AT, a mass-storage subsystem with a disk-storage capacity twice the size of the database and a nine-track tape drive, and an HP LaserJet printer for proofing. The user-interface software requires an HP Vectra PC or IBM PC/AT, CGA, EGA or monochrome monitor, 4 Mbytes of hard-disk storage and a CD-ROM drive. Hewlett-Packard is licensing the database-building software to publishers for \$50,000 (US); the user-interface software can be licensed for \$500 (US).

Beckman Instruments has recently introduced its CALS PeakPro **chromatography software** for the DEC VAX family of computers (Reader Service No. 105). PeakPro provides for the acquisition and



Beckman's PeakPro software now runs on VAX computers.

analysis of chromatographic data from both gas and liquid chromatographs. Analogue outputs from the instruments are interfaced to the computer using Beckman Digimetry Instrument Couplers, which connect to RS232 ports on either a terminal multiplexer or a DEC terminal server communicating to the computer over Ethernet. The PeakPro VAX software has full-screen colour menus with help keys, and integrates completely with Beckman's CALS Lab Manager networking software. For security, methods are protected from unauthorized modification, and an event log is kept of all significant events on the system. A snapshot of the analytical method is stored with the chromatographic results for full documentation of the analysis.

Bioscan has just launched its Optimas **image measurement and analysis software**, which can capture electronic images from microscopes, television, CT scanners, MR imagers and ultrasound imagers for analysis and storage on an IBM PC/AT or PS/2 (Reader Service No. 106).



Optimas: image analysis from Bioscan.

The program allows users to enhance images with standard image processing tools, and then extract a wide variety of measurement data. Data can be converted to ASCII files compatible with statistical analysis, spreadsheet, database, and text processing software, and can be exported to other programs through dynamic links. The Optimas software is based on a user-friendly Microsoft Windows format, but can be programmed for custom applications using an internal operating language called Alice, based on the C programming language. Repetitive procedures can be repeated exactly using "macro" commands to ensure that data is collected with the same procedure for

each specimen. For presentation and publication, Optimas can send data and images to word processing or page composition programs, and directly to laser printers and typesetting machines. Optimas costs \$3,995 (US).

Computer chemistry

Perkin-Elmer has a new **simulated distillation software package** available for use with its 8000 series gas chromatographs (Reader Service No. 107). The £600 (UK) program is designed to produce boiling point distribution data for gasolines and distillates according to set standards. The integral data handling of the 8000 series GCs are used to carry out routine data handling calculations, while a PC running IBM PC-DOS or MS-DOS performs the simulated distillation calculations. Separate functions allow boiling point and response factor calibrations, sample analysis, system performance checks, the downloading of disk-stored methods for setting up the GC and data handling parameters, and the establishment of a database comprising physical constants pertinent to the samples under analysis.

ADVERTISEMENTS

Economical Modular Tissue Culture Incubator

- Safe for AIDS research • Reliable, for In-Vitro fertilization
- Versatile enough for thousands of lab uses

Easy to use - Just flush with gas mixture, seal and place at approp. temp. Perform multiple experiments simultaneously without cross contamination. Thousands in use worldwide.

For information contact: **Flow Laboratories, Ltd.** or **Billups-Rothenberg, Inc.** • P.O. Box 977
Del Mar, CA 92014-0977 USA • (619) 755-3309

Reader Service No.25

O₂/CO₂ micro-Respirometer
Measures gaseous O₂/CO₂ uptake or production with 0.2 µl/h accuracy
Far superior to the Warburg Apparatus

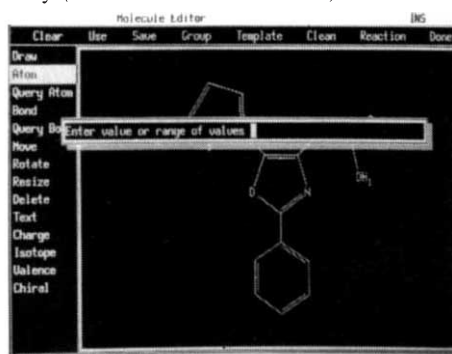
It can be used for measuring O₂/CO₂ production/consumption of insects, fish, small animals, plants, fruits, vegetables, bacterial and tissue cultures, fermentation, oxidation of metals and plastics etc. The results are printed periodically by IBM-PC/XT/AT.

Equipment is also available to measure oxygen consumption in larger animals (mice, rats, pigs, horses etc.).

COLUMBUS INSTRUMENTS
P.O. BOX 44049 COLUMBUS, OHIO 43204 U.S.A.
PHONE: (614) 488-6176 FAX: (614) 276-0529
TLX: 246514

Reader Service No.24

Molecular Design's latest version of its ChemBase **molecule and reaction database management system** for PCs has three new features which add to its flexibility (*Reader Service No. 108*). The new



ChemBase from Molecular Design.

atom aliasing feature allows scientists to attach nonstandard labels to an atom site without interfering with the ability of ChemBase to recognize the actual atom at the site. In the new version, users can also abbreviate any group to any text string — such as “Ph” for a phenyl group — and ChemBase will still recognize the underlying chemical structure. The software has been updated to allow atom-centred values, so that a chemist can indicate a value or range of values for individual atoms which can later be searched, a particularly useful characteristic for working with NMR data.

Hampden Data Services is now distributing STN Express **chemistry software** (*Reader Service No. 109*). STN Express gives chemists online access to a database of over nine million chemical structures, which can be searched from an IBM PC or compatible computer. STN Express has a graphic font, which simplifies search and retrieval by allowing chemists to draw the structure or sub-structure for which they want to search, using a mouse. Queries can be composed before connecting to the database to save connection time. Hampden Data Services also offers the PSIDOM software series for chemists to use to build up a personal graphical database of chemical structures on an IBM PC. PSIDOM interfaces with desktop publishing and word processing programs, and with STN Express. A demonstration disk including STN Express and the chemical structure drawing interface of PSIDOM is available from Hampden Data Services.

Networking

Access*LIMS, the **laboratory information management system** from Nelson Analytical designed for the DEC VAX and MicroVAX II computers, uses relational database technology (*Reader Service No. 110*). The Access*LIMS system can provide sample tracking, data collection and processing, *ad hoc* query capability, data validation, reporting, security and information archival. Nelson Analytical

says the system's relational data structure is flexible to use, and can evolve to cope with changes in the laboratory: as the nature of the data changes, so do the prompts, screens, forms and reports. Because a relational system uses automatic navigation to retrieve data, data can be found using a simple command, without the need to go through a series of menus. The Access*LIMS system can be upgraded to other DEC computers, and is fully compatible with other Nelson Analytical chromatography workstations and minicomputer-based systems. It can also be networked using Ethernet protocols. The Nelson Analytical first tier system includes the Access*LIMS software, DEC MicroVAX II minicomputer, terminal, printer and cabling for £60,000 (UK).

The Waters division of Millipore Corporation has a **laboratory networking computer system** that allows chromatography data to be sent directly to a VAX computer (*Reader Service No. 111*). The \$25,000–50,000 (US) Waters 860 networking computer system with the SAT/IN and LAC/E modules allows acquisition and transfer of real time chromatographic data directly to a VAX workstation or computer through a DECnet/Ethernet network for review, reprocessing or archiving. The SAT/IN modules acquire virtually noise-free data even at high sample rates, says Waters, for use in all types of LC, IC, GPC and capillary GC applications. Each LAC/E module receives input from up to four SAT/IN modules and transfers data using the DECnet/Ethernet network. The Waters 860 is based on the DEC MicroVAX line of processors, and performs chromatographic data reduction directly from a Waters data workstation or LAC/E module.

Dedicated instruments

Varian has just introduced its LC Star System, a fully integrated, modular **HPLC workstation** based on the IBM PS/2 computer (*Reader Service No. 112*). Each of the \$25,000–55,000 (US) LC Star System's modules — gradient pump, programmable autosampler, diode array detector, and UV-visible detector — can operate as stand-alone intelligent components without computer control.



HPLC on the IBM PS/2 system from Varian

But the Star 9025 software ties the modules together through the IBM PS/2 to allow the operating conditions and parameters for all components to be built into an automated method. The Windows environment of the IBM PS/2 helps the user to keep an eye on the entire process: one window can display the status of the operating LC components, while another can be opened to review results or to create methods to be used later. Yet another window can display colour graphic representations of system elements, such as the autosampler carousel.

An integral **data analysis and chemical information system** is included in VG Analytical's newest mass spectrometer, the AutoSpec (*Reader Service No. 113*).



The brains of VG Analytical's AutoSpec.

The AutoSpec is a trisector geometry (ESA-magnet-ESA) double-focusing instrument with a graphics-oriented computer control interface. The instrument can be changed from one ionization mode to another using “plug-in” sources. The VG Analytical SIOS interface between the mass spectrometer and its computer system is capable of making intelligent, data-dependent decisions, and the chemical information processing software is unique to the system.

The new delta gas chromatograph/mass spectrometer system from Finnigan MAT is controlled by the first software to provide real-time, **multitasking capabilities for isotope instruments**, says the company (*Reader Service No. 114*). The delta GC/MS instrument is the result of the direct interfacing of a GC to an isotope ratio MS through a combustion furnace. The delta GC converts organic compound components separated by the GC into CO₂, which is then purified without cryogenic coolants and directly introduced into the MS ion source. Finnigan MAT says the delta GC/MS provides measurement of carbon isotope content to better than 3.5×10^{-4} atom per cent excess. The ISODAT software which controls the instrument enables the updating of control tables and the evaluation of data during data acquisition. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.